

Towards optimal cosmological parameter recovery from compressed bispectrum statistics

Article (Published Version)

Byun, Joyce, Eggemeier, Alexander, Regan, Donough, Seery, David and Smith, Robert (2017) Towards optimal cosmological parameter recovery from compressed bispectrum statistics. Monthly Notices of the Royal Astronomical Society, 471 (2). pp. 1581-1618. ISSN 0035-8711

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/69105/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Towards optimal cosmological parameter recovery from compressed bispectrum statistics

Joyce Byun,^{*} Alexander Eggemeier,^{*} Donough Regan,^{*} David Seery
and Robert E. Smith

Astronomy Centre, School of Mathematical and Physical Sciences, University of Sussex, Brighton BN1 9QH, United Kingdom

Accepted 2017 June 30. Received 2017 June 27; in original form 2017 May 25

ABSTRACT

Over the next decade, improvements in cosmological parameter constraints will be driven by surveys of a large-scale structure in the Universe. The information they contain can be measured by suitably chosen correlation functions, and the non-linearity of structure formation implies that significant information will be carried by the 3-point function or higher correlators. Extracting this information is extremely challenging, requiring accurate modelling and significant computational resources to estimate the covariance matrix describing correlation between different Fourier configurations. We investigate whether it is possible to reduce this matrix without significant loss of information by using a proxy that aggregates the bispectrum over a subset of configurations. Specifically, we study constraints on Λ CDM parameters from a future galaxy survey combining the power spectrum with (a) the integrated bispectrum, (b) the line correlation function and (c) the modal decomposition of the bispectrum. We include a simple estimate for the degradation of the bispectrum with shot noise. Our results demonstrate that the modal bispectrum has comparable performance to the Fourier bispectrum, even using considerably fewer modes than Fourier configurations. The line correlation function has good performance, but is less effective. The integrated bispectrum is comparatively insensitive to the background cosmology. Addition of bispectrum data can improve constraints on bias parameters and σ_8 by a factor between 3 and 5 compared to power spectrum measurements alone. For other parameters, improvements of up to ~ 20 per cent are possible. Finally, we use a range of theoretical models to explore the sophistication required to produce realistic predictions for each proxy.

Key words: large-scale structure of Universe – cosmology: theory.

1 INTRODUCTION

Constraints on cosmological parameters have improved significantly over the last two decades, driven by high-precision data from the cosmic microwave background (‘CMB’) temperature and polarization anisotropies (Bennett et al. 2003; Ade et al. 2014). But the capacity of CMB observations to sustain this rate of progress is now nearly exhausted. Measurements of the temperature anisotropy have become limited by cosmic variance down to very small scales, and therefore future large-scale measurements will furnish little new information. Meanwhile, on small scales, cosmological information begins to be erased by astrophysical processes. Modest improvements may still come from better polarization data, perhaps shrinking current uncertainties by a factor of a few, but eventually these measurements will also approach the limit of cosmic variance. Further progress will be possible only with new sources of information. In the decade 2020–2030 we expect such a source to be provided by surveys of a cosmological large-scale structure – but only if the information these surveys contain can be extracted and understood (Silk 2017).

^{*} E-mail: joyce.byun@sussex.ac.uk (JB); a.eggemeier@sussex.ac.uk (AE); d.regan@sussex.ac.uk (DR)

1.1 The bispectrum: challenges

The statistical information contained in a galaxy survey is carried by the probability density of the fluctuations, which can be measured using a hierarchy of suitably chosen correlation functions, perhaps involving a transformation of the density field (Neyrinck, Szapudi & Szalay 2009; Seo et al. 2011; Carron & Szapudi 2013). Tools to extract information from the 2-point function were developed early and are now mature. The development of tools to extract information from higher-order correlation functions has proceeded more slowly (Fry 1984; Goroff et al. 1986; Scoccimarro 2000; Sefusatti et al. 2006), but because structure formation is non-linear, it is likely that these carry an important fraction of the information content. To make good use of our investment in costly observational programmes, it will be necessary to find a means of using information from at least the 3-point function. So far, there are few bispectrum (or 3-point function) measurements from modern surveys compared to power spectrum analyses (Marin et al. 2013; Gil-Marín et al. 2015, 2017; Slepian et al. 2015, 2017).

What are the challenges? A first difficulty arises from combinatorics. We write the matter overdensity at time t as $\delta(\mathbf{x}, t) = \delta\rho(\mathbf{x}, t)/\bar{\rho}(t)$, where $\delta\rho(\mathbf{x}, t) = \rho(\mathbf{x}, t) - \bar{\rho}(t)$ is the density perturbation and $\bar{\rho}(t)$ is the uniform background. Allowing angle brackets $\langle \dots \rangle$ to denote an ensemble average, statistical homogeneity makes the 2- and 3-point functions $\langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle$ and $\langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}_1)\delta(\mathbf{x} + \mathbf{r}_2) \rangle$ independent of the origin \mathbf{x} . After translation to Fourier space, this enforces conservation of momentum for the wavenumbers that participate in the expectation value,

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P(k), \quad (1a)$$

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(k_1, k_2, k_3), \quad (1b)$$

where $k = |\mathbf{k}_1| = |\mathbf{k}_2|$ is the common magnitude of the wavenumbers appearing in the 2-point function. In equations (1a)–(1b) and the remainder of this paper we suppress the time t labelling the hypersurface of evaluation. Isotropy makes the power spectrum P a function only of k , while the bispectrum B is a function of the three wavenumbers k_1, k_2, k_3 subject to the closure condition $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$. Therefore a fixed volume of space yields many more distinct configurations of the bispectrum than of the spectrum. If we choose to measure all of them then we must provide an estimate for their covariance, and beyond the Gaussian approximation this typically requires N -body simulations. Since we require at least as many simulations as the number of independent covariances, the number of simulations to be performed grows at least linearly in the number of configurations. This makes it very expensive to use more than a fraction of the available bispectrum measurements.

Second, we must estimate typical values for $B(k_1, k_2, k_3)$ in a particular cosmological model. While such estimates are already necessary for the power spectrum $P(k)$, accurate estimates for the bispectrum are substantially more challenging. There are two key reasons. No matter what methods we use, the algebraic complexity associated with high-order correlation functions is usually worse than at lower order. Also, many of our standard tools have a reduced range of validity as we move up the correlation hierarchy. We must therefore work harder to obtain trustworthy predictions from our models, and in some cases we can do so only by giving up analytic methods altogether.

These problems have hampered the development of a toolkit that would make use of bispectrum measurements routine. Nevertheless, they are difficulties of practice and not obstructions of principle – if necessary, we could determine both covariances and typical values of P or B from N -body simulations, at least over a certain range of scales. But such determinations would require a very large number of realizations. The sheer computational resource entailed by this strategy makes it unattractive on time-scales of interest for surveys such as Euclid, Dark Energy Spectroscopic Instrument (DESI) or Large Synoptic Survey Telescope (LSST).

1.2 Alternative strategies

To build a practical methodology, we must cut the size of the covariance matrices and avoid simulations where possible. Simulations are not needed when analytic methods suffice to predict P or B , or when a Gaussian approximation to the covariance is acceptable. Meanwhile, an obvious way to reduce the number of configurations is simply not to measure them all. Depending how aggressively we choose to cut, this may mean accepting a significant loss of information. A more nuanced option is to aggregate groups of configurations into weighted averages, effectively *compressing* the data carried by the bispectrum rather than discarding it. Such averages could be computed directly. But there are also observables whose statistics can naturally be expressed as weighted averages of this kind. Measuring these will often be simpler than measuring amplitudes of the Fourier bispectrum – simultaneously reducing the effort required to estimate and invert their covariance matrices. We describe these observables as ‘proxies’ or ‘proxy statistics’ for the full Fourier bispectrum.

Each proxy represents a compromise between (a) information loss due to compression, (b) the type of Fourier configurations over which it aggregates, and therefore the physics to which it is sensitive, and (c) its accessibility to analytical modelling, either for covariances or to estimate typical measurements. In this paper we select three proxies that have already been described in the literature and characterize their performance in each of these categories. Our aim is not to find an optimal proxy for any particular measurement, but rather to demonstrate that their use represents a feasible strategy for upcoming surveys without unacceptable degradation in information recovery.

1.3 Summary

Our principal results are forecasts for the parameter error bars achievable from combinations of the galaxy power spectrum and bispectrum, or its proxies. The parameter set we study comprises the background quantities of a Λ CDM model with evolving dark energy, supplemented by two parameters describing the bias model (McDonald & Roy 2009). We study how these forecasts change when they are estimated using the complete non-Gaussian covariance matrix or its Gaussian approximation. We characterize their dependence on the method used to predict typical values for $P(k)$ and $B(k_1, k_2, k_3)$ by sampling the results using tree-level and one-loop standard perturbation theory ('SPT'), and an implementation of the halo model. We compare these estimates with values measured directly from simulations. These results can be used to determine, for each observable, the degree of modelling sophistication that is required to obtain accurate forecasts.

Our analysis does not include the effect of survey geometry or incompleteness, or redshift-space effects, and should be regarded as a determination of the performance of each proxy under idealized conditions. We include a simple analysis that indicates how our results would change in the presence of shot noise.

Fisher forecasts, including Fourier bispectrum measurements, have previously been reported by Sefusatti et al. (2006), assuming 1015 bispectrum configurations and measuring covariances from a suite of 6000 mock catalogues generated by the PTHalos algorithm (Scoccimarro & Sheth 2002) and second-order Lagrangian perturbation theory ('2LPT'). Their results suggested that the bispectrum contains significant cosmological information. For comparison, in our analysis we use 95 bispectrum configurations in order to keep the size of the covariance matrix within plausible bounds, and measure it directly from a suite of full N -body simulations.

More recently, Chan & Blot (2016) estimated the extra constraining power of Fourier bispectrum measurements by computing their contribution to the signal to noise, but did not make forecasts for error bars on cosmological parameters. They concluded that the bispectrum contributes a modest increase in the signal to noise above the power spectrum, with the additional information perhaps being principally useful to break degeneracies. One of our aims is to clarify the relationship between this conclusion and the more nuanced outcomes found by Sefusatti et al. (2006). We find that estimates based on signal to noise alone generally give only a rough indication compared to the full Fisher calculation because they do not account for variations in the sensitivity to background cosmology between observables.

1.4 Organization

Our presentation is organized as follows. In Section 2 we introduce the three bispectrum proxies to be studied in the remainder of the paper. These are: (a) the *integrated bispectrum* (Chiang et al. 2014), which measures variation of the power spectrum in subsampled regions; (b) the *line correlation function*, which samples 3-point statistics of the phase of the density fluctuation (Obreschkow et al. 2013; Wolstenhulme, Bonvin & Obreschkow 2015), and (c) the *modal bispectrum*, which can be regarded as an alternative to the Fourier bispectrum obtained by exchanging the Fourier modes $e^{ik \cdot x}$ for an alternative basis (Fergusson, Regan & Shellard 2012; Regan et al. 2012). Each of these measures can be expressed as a weighted average over particular configurations of the Fourier bispectrum.

In Sections 3.2–3.4 we explain how each proxy can be predicted using the halo model or a flavour of SPT. In Section 3.5 we explain our prescription to obtain the biased galaxy density field from the underlying matter density field, which is the quantity predicted by these analytic models. In Section 4 we describe our procedure to recover estimates for each proxy statistic from N -body simulations, and in Section 5 we compare these estimates (and estimates for their derivatives with respect to the cosmological parameters) with theoretical predictions. Readers familiar with the measures of 3-point correlations described in Section 2 and the modelling technologies of Section 3 may choose to begin reading at this point. In Section 6 we present signal-to-noise estimates for the information content of each proxy. Our Fisher forecasts appear in Section 7. In Section 8 we collect a number of topics for discussion, including the compression efficiency of each proxy statistic and the impact of shot noise on our forecasts. We conclude in Section 9.

1.5 Notation

Our Fourier convention is $f(\mathbf{x}) = \int d^3k (2\pi)^{-3} f(\mathbf{k}) e^{ik \cdot x}$. To avoid confusion, we distinguish the Dirac δ -function $\delta_D(\mathbf{x})$ or $\delta_D(\mathbf{k})$ and the Kronecker symbol $\mathbf{1}_{ij}$ from the matter overdensity $\delta \equiv \delta\rho/\rho$.

2 THE FOURIER BISPECTRUM AND ITS PROXIES

In this section we introduce the proxy statistics to which we compare the Fourier bispectrum. This has already been defined – together with the power spectrum – in equations (1a)–(1b). We describe the integrated bispectrum in Section 2.1, the line correlation function in Section 2.2 and the modal decomposition of the bispectrum in Section 2.3. Each of these represents a possible compression of the Fourier bispectrum, in the sense described in Section 1.

2.1 Integrated bispectrum

The integrated bispectrum (or 'position-dependent power spectrum') was developed by Chiang et al. (2014) as a tool to search for primordial non-Gaussianity in a large-scale structure. It has several convenient features: it is easily estimated using standard power-spectrum codes and it has a clear physical interpretation. As we shall see in Section 3.2, it represents a weighted average of the Fourier bispectrum dominated

by ‘squeezed’ configurations – that is, wavenumbers $(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ where one k_i is much smaller than the other two. If we assume $k_3 \ll k_1, k_2$, then the bispectrum $\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle$ expresses correlations between a single long-wavelength mode $\delta(\mathbf{k}_3)$ and the 2-point function $\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle$. This makes it sensitive to ‘local-type’ non-Gaussianity produced by inflationary models with more than one active field. However, because gravitational collapse correlates modes with comparable wavenumbers, the bispectrum produced during mass assembly is typically concentrated away from squeezed configurations. For this reason it is not clear how sensitive the integrated bispectrum might be to the cosmological parameters that influence this assembly process.

To define the integrated bispectrum, divide the total survey volume into N_s cubic subvolumes, each of volume $V_s \equiv L_s^3$ and centred at positions \mathbf{r}_L . Compute the power spectrum and average overdensity for each subvolume, which we denote $P(\mathbf{k}, \mathbf{r}_L)$ and $\bar{\delta}(\mathbf{r}_L)$, respectively. [The power spectrum $P(\mathbf{k}, \mathbf{r}_L)$ may depend on the orientation of \mathbf{k} if the subvolumes are not isotropic.] Finally, the integrated bispectrum is defined to be the expectation of $P(\mathbf{k}, \mathbf{r}_L)\bar{\delta}(\mathbf{r}_L)$, averaged over the orientation of \mathbf{k} ,

$$iB(k) \equiv \int \frac{d^2\hat{k}}{4\pi} \langle P(\mathbf{k}, \mathbf{r}_L)\bar{\delta}(\mathbf{r}_L) \rangle_{N_s}. \quad (2)$$

The notation $\langle \cdots \rangle_{N_s}$ indicates that the expectation is to be taken over all subvolumes.

To compute this expectation, we Taylor expand $P(\mathbf{k}, \mathbf{r}_L)$ in powers of $\bar{\delta}(\mathbf{r}_L)$ (Chiang et al. 2014). The leading contribution is

$$\langle P(\mathbf{k}, \mathbf{r}_L)\bar{\delta}(\mathbf{r}_L) \rangle_{N_s} = \left\langle \left[P(\mathbf{k}) \Big|_{\bar{\delta}=0} + \frac{dP(\mathbf{k})}{d\bar{\delta}} \Big|_{\bar{\delta}=0} \bar{\delta}(\mathbf{r}_L) + \cdots \right] \bar{\delta}(\mathbf{r}_L) \right\rangle_{N_s} \approx \frac{d \ln P(\mathbf{k})}{d\bar{\delta}} \Big|_{\bar{\delta}=0} P(\mathbf{k}) \sigma_L^2, \quad (3)$$

where $\sigma_L^2 \equiv \langle \bar{\delta}^2(\mathbf{r}_L) \rangle_{N_s}$ is the variance in mean overdensity over the subvolumes. Therefore, at the lowest order, the integrated bispectrum describes variation of the power spectrum in response to changes in the large-scale overdensity.¹ We conclude that measurements of iB contain both the power spectrum and its variance. Since these can be measured directly, any new information contained in the integrated bispectrum must reside in its normalized component (Chiang et al. 2014),

$$ib(k) \equiv \frac{iB(k)}{P(k)\sigma_L^2} \approx \frac{d \ln P(k)}{d\bar{\delta}} \Big|_{\bar{\delta}=0}, \quad (4)$$

where the second approximate equality applies when only the lowest-order contribution from the Taylor expansion need to be retained. This is the linear response approximation. The quantity $d \ln P(k)/d\bar{\delta}$ is the *linear response function* and provides a good approximation to ib for large k .

2.2 Line correlation function

Equation (1a) shows that the power spectrum is sensitive only to information carried by the amplitude of each Fourier mode. In contrast, higher-order statistics generally encode information carried by both amplitudes and phases. Phase correlations are an exclusive signature of non-Gaussian density fields. For instance, they may arise through processes in the primordial Universe or from mode coupling in the non-linear regime of gravitational collapse. Therefore, unlike the amplitudes, phases directly probe cosmological information that is absent from the 2-point function.

With this motivation, Obreschkow et al. (2013) proposed the line correlation function (often abbreviated as ‘LCF’). It measures a subset of 3-point phase correlations of the density field – specifically, correlations between collinear points, each separated by a distance r . Obreschkow et al. (2013) demonstrated that the LCF is a robust tracer of filamentary structures, and showed that it could be used as a phenomenological tool to distinguish between cold and warm dark matter scenarios. Subsequent work established its connection to conventional higher-order statistics (Eggemeier et al. 2015; Wolstenhulme et al. 2015; Eggemeier & Smith 2017).

The line correlation function can be understood as follows: for a given density field $\delta(\mathbf{x})$ in some volume V , its real-space phase field $\epsilon_r(\mathbf{x})$ smoothed on a scale r satisfies

$$\epsilon_r(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \epsilon(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}} W(k|r) \equiv \int \frac{d^3k}{(2\pi)^3} \frac{\delta(\mathbf{k})}{|\delta(\mathbf{k})|} e^{i\mathbf{k} \cdot \mathbf{x}} W(k|r), \quad (5)$$

where $w(k|r)$ is the Fourier transform of the smoothing window function. We take this to be a spherical top-hat in k -space, $W(k|r) \equiv \Theta(1 - kr/2\pi)$, where $\Theta(x)$ denotes the Heaviside step function. The phase at $\mathbf{k} = 0$ is defined so that $\epsilon(\mathbf{0}) \equiv 0$. Following Obreschkow et al. (2013), the LCF is defined by

$$\ell(r) \equiv \frac{V^3}{(2\pi)^9} \left(\frac{r^3}{V} \right)^{3/2} \int \frac{d^2\hat{r}}{4\pi} \langle \epsilon_r(\mathbf{x}) \epsilon_r(\mathbf{x} + \mathbf{r}) \epsilon_r(\mathbf{x} - \mathbf{r}) \rangle, \quad (6)$$

where the factor $V^3/(2\pi)^9$ represents a volume regularization. After taking Fourier transforms, we require the 3-point function of the $\epsilon_r(\mathbf{k})$ in order to evaluate this integral. Wolstenhulme et al. (2015) and Eggemeier & Smith (2017) demonstrated that, at the lowest order in the expansion of the probability density function for Fourier phases, this 3-point function is directly related to the Fourier bispectrum. Therefore

¹ In field theory this is the ‘operator product expansion’.

the LCF must contain some fraction of the information in B , but because $\ell(r)$ is an average over specific collinear configurations, it represents a compression. Specifically, the number of LCF bins will vary linearly with changes in the effective cut-off on Fourier modes.

2.3 Modal bispectrum

Our final proxy is a ‘modal’ expansion of the 3-point function. This is very similar to the Fourier bispectrum, except that we exchange the Fourier basis $e^{ix \cdot k}$ for a set of alternative modes that are better adapted to the structure of B . The exchange is helpful if we can represent the bispectrum to the same accuracy using fewer modes than required by the Fourier representation. This approach was originally developed by Fergusson & Shellard (2009) and Regan, Shellard & Fergusson (2010) to analyse microwave background data, and subsequently applied to the large-scale structure by Fergusson et al. (2012) and Regan et al. (2012).

In the alternative basis we represent the Fourier bispectrum in the form

$$B(k_1, k_2, k_3) \approx B_{\text{modal}}(k_1, k_2, k_3) \equiv \frac{1}{w(k_1, k_2, k_3)} \sum_{n=0}^{n_{\text{max}}-1} \beta_n^Q Q_n(k_1, k_2, k_3), \quad (7)$$

where the Q_n are basis functions that span the space of configurations compatible with a triangle condition on (k_1, k_2, k_3) , but can otherwise be chosen freely provided they are linearly independent. The β_n^Q are numbers that we describe as ‘modal coefficients’. They can be regarded as averages of the Fourier bispectrum over a set of configurations picked out by the corresponding Q_n . The function $w(k_1, k_2, k_3)$ is an arbitrary weight that will be chosen in Section 3.3.

If the Q_n form a complete basis we expect B and B_{modal} to become equivalent in the limit $n_{\text{max}} \rightarrow \infty$. In this limit the modal expansion is merely a reorganization of the Fourier representation. But if we select the lowest Q_n to average over the most relevant Fourier configurations, then it may be possible to represent a typical B using only a small number of modes.² Taking n_{max} to be of order this number, the outcome yields useful compression whenever $n_{\text{max}} \ll N_{\text{triangles}}$, where $N_{\text{triangles}}$ is the number of Fourier configurations contained in the volume under discussion. At least for reasonably smooth bispectra, Schmittfull, Regan & Shellard (2013) found that this could be done with no more than modest loss of signal.

2.3.1 Orthonormal basis

Given a choice of Q_n we may redefine the basis by taking arbitrary linear combinations. For example, we will use this freedom in Section 3.4 to obtain a basis for which the β -coefficients are uncorrelated. The covariance matrix in this redefined basis is especially simple.

Such a redefinition can be performed using an invertible matrix λ_{mn} . We define $R_n \equiv \sum_m \lambda_{nm}^{-1} Q_m$. The β -coefficients in the R basis now satisfy $\beta_n^R \equiv \sum_m \lambda_{mn} \beta_m^Q$. Since the Q - and R -bases are reorganizations of each other, the modal bispectrum defined using either basis is equivalent,

$$B(k_1, k_2, k_3) \approx \frac{1}{w(k_1, k_2, k_3)} \sum_{n=0}^{n_{\text{max}}-1} \beta_n^Q Q_n(k_1, k_2, k_3) = \frac{1}{w(k_1, k_2, k_3)} \sum_{n=0}^{n_{\text{max}}-1} \beta_n^R R_n(k_1, k_2, k_3). \quad (8)$$

3 PREDICTING TYPICAL VALUES AND COVARIANCES FOR THE PROXIES

In this section we explain how to obtain predictions for the typical values and covariances of $ib(k)$, $\ell(r)$ and β_m^R in a given cosmological model. This can be done with different degrees of sophistication, corresponding – for example – to truncations at different levels in the loop expansion of standard perturbation theory (Bernardeau et al. 2002), or by using fitting functions calibrated to match the output of N -body simulations (Mead et al. 2015). Since each proxy aggregates a different group of Fourier configurations, and these configurations vary in their response to features of the background cosmology, the sophistication needed to adequately capture the behaviour of the proxies may vary.

This is both a challenge and an opportunity. Proxies that require delicate modelling to obtain accurate predictions are harder to use, and may be expensive to deploy in a parameter-estimation Monte Carlo. In favourable cases, however, the payoff will be sensitive discrimination between nearby cosmological models. On the other hand, proxies that can be modelled robustly using simple methods are easy to use and cheap to deploy, but may offer correspondingly coarse discrimination. We study these trade-offs by contrasting predictions made using tree-level and one-loop SPT, and the halo model. For the halo-model power spectrum we choose the HMcode implementation (Mead et al. 2015). For the halo-model bispectrum we use the standard formulae given by Cooray & Sheth (2002) with a Sheth–Tormen mass function (Sheth & Tormen 1999) and Navarro–Frenk–White halo profile (Navarro, Frenk & White 1996). In Section 5 we study the performance of each method compared to numerical estimates extracted directly from N -body simulations, which enables us to characterize the minimum adequate sophistication for each proxy. For simplicity our analysis is framed in terms of the underlying dark matter density field, although in Section 3.5 we explain how this can be extended to predict galaxy clustering.

² Here, ‘most relevant’ is defined by the features of the bispectrum for which we wish to search. For example, inspection of the formulae appearing in Sections 3.2–3.3 below shows that both the integrated bispectrum and line correlation function can be regarded as instances of (7), with Q_n adjusted to prioritize specific groups of Fourier configurations. For these cases, however, the resulting Q basis is not complete. In this paper we distinguish the modal decomposition, for which the Q basis is intended to be complete, from proxies such as ib and ℓ , which are intended to be projections.

3.1 Covariance

To compute a likelihood for a given proxy, either for the purposes of parameter estimation or to make forecasts, we require an estimate for the covariance between different configurations. Therefore the minimum sophistication needed to adequately predict this covariance matrix will play an additional role in determining the relative expense of each proxy. In practice the covariance matrix is typically estimated by taking measurements from a large suite of N -body simulations or 2LPT catalogues, or, if this is cannot be done, by falling back to a Gaussian approximation. N -body simulations give accurate results, but are expensive enough that assembling sufficient independent realizations to determine the inverse covariance is often not feasible. In comparison, catalogues based on 2LPT are significantly cheaper but become inaccurate in the non-linear regime, while the Gaussian prediction breaks down even earlier and may miss cross-correlations that significantly affect the outcome.

The relative importance of these cross-correlations varies between proxies. In Sections 6–7 we estimate their significance by comparing results from N -body and Gaussian covariances. We describe our procedure to estimate covariance matrices from the simulations in Section 5, but collect formulae for the Gaussian approximation here.

For comparison, the Gaussian covariance for the power spectrum and Fourier bispectrum, measured on a grid of spacing Δk with fundamental frequency $k_f = 2\pi/V^{1/3}$, can be written

$$\text{Cov}_G[P(k_i), P(k_j)] \approx \mathbf{1}_{ij} \frac{2k_f^3}{4\pi k_i^2 \Delta k} P^2(k_i), \quad (9)$$

where $\mathbf{1}_{ij}$ is the Kronecker symbol, and

$$\text{Cov}_G[B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), B(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)] \approx \mathbf{1}_{\mathbf{k}, \mathbf{q}} \frac{\mathbf{N} \pi k_f^3}{k_1 k_2 k_3 (\Delta k)^3} P(k_1) P(k_2) P(k_3). \quad (10)$$

The Kronecker symbol $\mathbf{1}_{\mathbf{k}, \mathbf{q}}$ should be interpreted to equal unity if the triangles defined by $\{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3\}$ and $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ are equal, and zero otherwise. The degeneracy factor \mathbf{N} equals unity for a scalene triangle, two for an isosceles triangle and six for an equilateral triangle.

3.2 Integrated bispectrum

To evaluate expression (4), we first establish its relation to the underlying 3-point function. The overdensity within the subvolume labelled by \mathbf{r}_L can be written

$$\delta(\mathbf{k}, \mathbf{r}_L) = \int \frac{d^3 q}{(2\pi)^3} \delta(\mathbf{k} - \mathbf{q}) W_L(\mathbf{q}) e^{-i\mathbf{q} \cdot \mathbf{r}_L}, \quad (11)$$

where $W_L(\mathbf{q}) = V_s \prod_{i=1}^3 \text{sinc}(q_i L_s/2)$ is the Fourier transform of the cubic window function with side length L_s , and $\text{sinc} x \equiv (\sin x)/x$. The power spectrum in this subvolume is $P(\mathbf{k}, \mathbf{r}_L) \equiv \langle |\delta(\mathbf{k}, \mathbf{r}_L)|^2 \rangle / V_s$ and the mean overdensity is $\bar{\delta}(\mathbf{r}_L) \equiv \delta(\mathbf{0}, \mathbf{r}_L) / V_s$. Combining these with equation (2) yields (Chiang et al. 2014)

$$i B^{\text{theory}}(\mathbf{k}) = \frac{1}{V_s^2} \int \frac{d^2 \hat{k}}{4\pi} \int \frac{d^3 q_1}{(2\pi)^3} \int \frac{d^3 q_2}{(2\pi)^3} B^{\text{theory}}(\mathbf{k} - \mathbf{q}_1, -\mathbf{k} + \mathbf{q}_1 + \mathbf{q}_2, -\mathbf{q}_2) W_L(\mathbf{q}_1) W_L(-\mathbf{q}_1 - \mathbf{q}_2) W_L(\mathbf{q}_2). \quad (12)$$

Because $\text{sinc} x$ is strongly peaked for $|x| \lesssim \pi$ the window functions W_L effectively constrain the q_i integrals to $q_i \lesssim 1/L_s$. Since $k \gtrsim 1/L_s$ within each subvolume, the integral receives significant contributions only from squeezed configurations of the Fourier bispectrum that are of order the subvolume size or larger, because in the limit $q_1, q_2 \ll k$ we have $B^{\text{theory}}(\mathbf{k} - \mathbf{q}_1, -\mathbf{k} + \mathbf{q}_1 + \mathbf{q}_2, -\mathbf{q}_2) \approx B^{\text{theory}}(\mathbf{k}, -\mathbf{k}, -\mathbf{q}_2)$.

Chiang et al. (2014) computed the linear response function using (12) and tree-level SPT, and verified that it reproduces equation (4) to within 2 per cent for $k \gtrsim 0.2 h^{-1} \text{ Mpc}$. For our purposes we require accurate estimates at smaller k , and therefore we perform a numerical integration using (12) directly. The integral is 8-dimensional and its evaluation is challenging; we implement it using the *Vegas* algorithm provided by the *CUBA* package (Hahn 2016). To make the integration time feasible, we densely sample B^{theory} on a 3-dimensional cubic mesh in coordinates (k_1, k_2, μ_{12}) , where $\mu_{12} \equiv (k_1^2 + k_2^2 - k_3^2)/(2k_1 k_2)$ is the cosine of the angle between \mathbf{k}_1 and \mathbf{k}_2 and can be used in place of the third wavenumber k_3 . We construct a 3-dimensional cubic spline that interpolates between lattice points and use this spline to evaluate the integrand. To validate this procedure, we have verified that our numerical results match the analytic prediction from the linear response function at large k .

Although we have not written subvolume labels explicitly, σ_L^2 and all power spectra in (4) refer to subsampled quantities, and therefore should be computed by appropriate convolution with the subvolume window function $W_L(\mathbf{q})$.

3.2.1 Halo model

This procedure yields good results for tree-level and one-loop SPT, but does not perform well when applied to the halo model. In this case we do not recover equivalence between our evaluation of (12) and the linear response function, which we compute by numerical differentiation of the *HMcode* power spectrum. The difference between (12) and the linear response function is 3–7 per cent for $k \gtrsim 0.2 h^{-1} \text{ Mpc}$. We interpret this disagreement as an indication that the standard halo model makes inconsistent predictions for the modulation of the power spectrum with $\bar{\delta}$, or the squeezed limit of the bispectrum or both. Moreover, comparison of the halo-model ib computed using (12) to our N -body simulations

shows poor agreement, suggesting that estimates based on (12) will be inaccurate. Therefore, for the halo model only, we estimate ib by assuming the linear response approximation (4) and computing $\ln P / d\delta$. We calculate the derivative using the simulation-calibrated formula proposed by Chiang et al. (2014),

$$\frac{d \ln P^{\text{halo}}(k)}{d\delta} = \frac{13}{21} \frac{d \ln P^{\text{halo}}(k)}{d \ln \sigma_8} + 2 - \frac{1}{3} \frac{d \ln k^3 P^{\text{halo}}(k)}{d \ln k}, \quad (13)$$

which agrees with our simulations to within 2 per cent for $k \gtrsim 0.2 h^{-1} \text{ Mpc}$.

3.2.2 Covariance

In the absence of shot noise, the Gaussian covariance for estimates of ib constructed from data can be written

$$\text{Cov}_G[ib(k_i), ib(k_j)] = \frac{V_s}{V N_{ks}} \frac{1}{\sigma_L^2} \mathbf{1}_{ij}. \quad (14)$$

In this expression V_s is the volume of a subsampled region and V denotes the total survey volume. The quantity $N_{ks} = 2\pi k^2 \Delta k V_s$ is the number of Fourier modes in a subvolume k -bin.

3.3 Line correlation function

Wolstenhulme et al. (2015) used tree-level SPT to predict the line correlation function. Their result was generalized to an arbitrary bispectrum by Eggemeier & Smith (2017), who gave the formula

$$\ell^{\text{theory}}(r) \simeq \left(\frac{r}{4\pi}\right)^{9/2} \iint_{\substack{|\mathbf{k}_1|, |\mathbf{k}_2| \\ |\mathbf{k}_1 + \mathbf{k}_2| \leq 2\pi/r}} d^3 k_1 d^3 k_2 B_\epsilon^{\text{theory}}(k_1, k_2, k_3) j_0(|\mathbf{k}_1 - \mathbf{k}_2| r), \quad (15)$$

where $j_0(x) = \sin(x)/x$ is the spherical Bessel function of order zero and the integrals over \mathbf{k}_1 and \mathbf{k}_2 are cut off at the scale $k_i = 2\pi/r$. The quantity B_ϵ is defined by

$$B_\epsilon(k_1, k_2, k_3) \equiv \frac{B(k_1, k_2, k_3)}{\sqrt{P(k_1)P(k_2)P(k_3)}} \quad (16)$$

and gives the dominant contribution to the bispectrum of the phase field $\epsilon(\mathbf{k}) = \delta(\mathbf{k})/|\delta(\mathbf{k})|$ in the limit of large volume V ,

$$\langle \epsilon(\mathbf{k}_1) \epsilon(\mathbf{k}_2) \epsilon(\mathbf{k}_3) \rangle = \frac{(2\pi)^3}{V^{3/2}} \left(\frac{\sqrt{\pi}}{2}\right)^3 B_\epsilon(k_1, k_2, k_3) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3), \quad (17)$$

while for smaller volumes there are corrections scaling as powers of $V^{-1/2}$ compared to the dominant term (Eggemeier & Smith 2017). In the following we will therefore refer to B_ϵ as the *phase bispectrum*.

3.3.1 Evaluation

To evaluate (15), we must perform a 6-dimensional integral. We use a strategy similar to that described in Section 3.2, by sampling the bispectrum over a cubic lattice and interpolating between lattice sites. The integration is again performed using Vegas.

In the special case of tree-level SPT, Wolstenhulme et al. (2015) showed that (15) could be reduced to a 3-dimensional integral,

$$\begin{aligned} \ell^{\text{tree}}(r) = 16\pi^2 \left(\frac{r}{4\pi}\right)^{9/2} \int_0^{2\pi/r} dk_1 k_1^2 \int_0^{2\pi/r} dk_2 k_2^2 \int_{-1}^{\mu_{\text{cut}}} d\mu F_2^{(s)}(k_1, k_2, \mu) \sqrt{\frac{P^{\text{tree}}(k_1)P^{\text{tree}}(k_2)}{P^{\text{tree}}(|\mathbf{k}_1 + \mathbf{k}_2|)}} \\ \times \left[j_0(|\mathbf{k}_2 - \mathbf{k}_1| r) + 2j_0(|\mathbf{k}_1 + 2\mathbf{k}_2| r) \right], \end{aligned} \quad (18)$$

where P^{tree} is the tree-level power spectrum, and

$$F_2^{(s)}(k_1, k_2, \mu) = \frac{5}{7} + \frac{\mu}{2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{2}{7} \mu^2 \quad (19)$$

is the symmetrized second-order SPT (Bernardeau et al. 2002) kernel with μ being the angle between \mathbf{k}_1 and \mathbf{k}_2 . The upper limit of the μ -integral is chosen to guarantee $|\mathbf{k}_1 + \mathbf{k}_2| \leq 2\pi/r$, which requires

$$\mu_{\text{cut}} = \min \left\{ 1, \max \left\{ -1, \frac{(2\pi/r)^2 - k_1^2 - k_2^2}{2k_1 k_2} \right\} \right\}. \quad (20)$$

Equation (18) is useful because it provides a means to test the accuracy of our 6-dimensional Vegas integrations, and the 3-dimensional interpolations they entail. We have compared estimates for the tree-level line correlation function using both (15) and (18) and find that the agreement is better than 0.1 per cent at the smallest r bin considered and worsens to only ~ 2 per cent at the largest one.

3.3.2 Covariance

To determine the Gaussian covariance, we require the 2-point function of the phase field,

$$\langle \epsilon(\mathbf{k}_1) \epsilon(\mathbf{k}_2) \rangle = \frac{(2\pi)^3}{V} \delta_D(\mathbf{k}_1 + \mathbf{k}_2). \quad (21)$$

It follows that, in the absence of shot noise, the covariance between estimators for the line correlation function on scales r_i and r_j can be written (Eggemeier & Smith 2017)

$$\text{Cov}_G[\ell(r_i), \ell(r_j)] = \frac{(r_i r_j)^{9/2}}{V^3} \iint_{\substack{|\mathbf{k}_1|, |\mathbf{k}_2| \\ |\mathbf{k}_1 + \mathbf{k}_2| \leq 2\pi/r}} \frac{d^3 k_1}{k_f^3} \frac{d^3 k_2}{k_f^3} \left(j_0(|2\mathbf{k}_1 + \mathbf{k}_2| r_i) [2j_0(|\mathbf{k}_1 - \mathbf{k}_2| r_j) + j_0(|2\mathbf{k}_1 + \mathbf{k}_2| r_j)] + r_i \leftrightarrow r_j \right), \quad (22)$$

where $k_f = 2\pi/V^{1/3}$ denotes the fundamental frequency (defined above equation (9)), and $r = \max\{r_i, r_j\}$. Note that (22) is not diagonal; the integral that defines the line correlation function depends on a range of Fourier modes for any scale r_i , and any Fourier modes that are common between $\ell(r_i)$ and $\ell(r_j)$ will contribute a nonzero covariance. Moreover, equation (22) shows that the Gaussian covariance is independent of redshift and all cosmological parameters.

3.4 Modal bispectrum

It was explained in Section 2.3 that the modal decomposition is defined by choice of a basis Q_n that samples groups of relevant Fourier configurations. The structure and ordering of the Q_n determine those configurations we wish to prioritize. But unless we carefully adjust the Q_n they will be correlated, and these correlations will be inherited by the β_n^Q . The outcome is that the covariance matrix for estimators of the β_n^Q is rather complex.

3.4.1 Construction of R basis

To avoid this, we redefine the basis, as in equation (8), to simplify the covariance matrix for estimators of the corresponding β_n^R . The construction proceeds in stages. First, consider the expected signal to noise with which it is possible to measure a single mode Q_n/w from (7). Using a Gaussian approximation for the noise, this can be written

$$6 \left(\frac{S}{N} \right)_{Q_n}^2 = \int \frac{d^3 k_1}{(2\pi)^3} \frac{d^3 k_2}{(2\pi)^3} \frac{d^3 k_3}{(2\pi)^3} (2\pi)^3 \frac{\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)}{w(k_1, k_2, k_3)^2} \frac{Q_n(k_1, k_2, k_3)^2}{P(k_1)P(k_2)P(k_3)}. \quad (23)$$

We are free to choose the weight w to simplify this integral. We define³

$$w(k_1, k_2, k_3) = \sqrt{\frac{k_1 k_2 k_3}{P(k_1)P(k_2)P(k_3)}}, \quad (24)$$

after which the computation of the expected signal to noise reduces to

$$6 \left(\frac{S}{N} \right)_{Q_n}^2 = \langle \langle Q_n | Q_n \rangle \rangle. \quad (25)$$

To write this and similar expressions economically, we have introduced the notation

$$\langle \langle f | g \rangle \rangle \equiv \int \frac{d^3 k_1}{(2\pi)^3} \frac{d^3 k_2}{(2\pi)^3} \frac{d^3 k_3}{(2\pi)^3} (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \frac{f(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{k_1 k_2 k_3} \quad (26)$$

for any f and g . In the special case that these depend only on the wavenumbers k_i and not their orientations $\hat{\mathbf{k}}_i$, some of the angular integrations are trivial and we obtain the simpler expression

$$\langle \langle f | g \rangle \rangle \equiv \frac{1}{8\pi^4} \int_{\mathcal{V}} dk_1 dk_2 dk_3 f(k_1, k_2, k_3) g(k_1, k_2, k_3). \quad (27)$$

Here, \mathcal{V} represents the set of points (k_1, k_2, k_3) where lines of length k_1, k_2 and k_3 can be arranged to form a triangle, ie. $2 \max\{k_i\} \leq \sum_i k_i$; for details, see Fergusson, Liguori & Shellard (2010). In principle, the integral can be carried over all k_i , but in practice it will be cut off at upper and lower limits k_{\max} and k_{\min} . Expressions (26) and (27) can be regarded as an inner product on the Q_n that weights each contributing Fourier configuration according to its individual signal-to-noise.

³ In fact for our estimation procedure we shall use the measured power spectrum \hat{P} in place of P in the formula for the weighting function. This makes little difference to the expectation value, but is preferred as it makes no knowledge about the underlying cosmology.

Second, the R basis is chosen to be diagonal with respect to this inner product. As we will see below, because the resulting R_n modes are orthogonal when weighted by signal to noise, the covariance matrix for estimators of the coefficients β_n^R becomes diagonal under the same approximation of Gaussian noise used to determine the weighting in (23). Specifically, we define

$$\langle\langle Q_m | Q_n \rangle\rangle \equiv \gamma_{mn} \equiv \frac{(k_{\max} - k_{\min})^3}{8\pi^4} \bar{\gamma}_{mn}. \quad (28)$$

It is sometimes preferable to express results in terms of $\bar{\gamma}_{mn}$, which is independent of k_{\min} and k_{\max} . For any suitable Q basis both γ_{mn} and $\bar{\gamma}_{mn}$ will be symmetric and positive-definite and may be factored into the product of a matrix and its transpose. Therefore there exists a matrix λ_{mn} such that $\bar{\gamma}_{mn} = \sum_r \lambda_{mr} \lambda_{nr}$. Application of (8) with λ_{mn} as the transformation matrix yields $R_n = \sum_{n'} \lambda_{nn'}^{-1} Q_{n'}$, and these modes are orthogonal in the sense

$$\langle\langle R_m | R_n \rangle\rangle = \frac{(k_{\max} - k_{\min})^3}{8\pi^4} \mathbf{1}_{mn}. \quad (29)$$

3.4.2 Determination of modal coefficients

Whether we work with the Q or R basis, we must predict the corresponding β -coefficients for each model of interest. In practice the extra matrix operations needed to obtain the R basis mean that it is simplest to perform calculations in the Q basis, before translating to the R basis to interpret the results. We adopt this procedure whenever concrete calculations using the modal decomposition are required. We use the Q basis constructed by Fergusson et al. (2010). (The details are summarized in Appendix A1.) It is not intended to prioritize any single class of Fourier configurations, but rather attempts to provide a good description of reasonably smooth bispectra over a range of shapes and scales.

To extract the β_n^Q we use (26). Assuming (7) can be interpreted as an equality, we conclude that for an arbitrary bispectrum $B^{\text{theory}}(k_1, k_2, k_3)$

$$\langle\langle w B^{\text{theory}} | Q_m \rangle\rangle = \sum_{n=0}^{n_{\max}-1} \beta_n^{Q, \text{theory}} \gamma_{nm}. \quad (30)$$

Finally, the individual β_n^Q should be extracted by contraction with the inverse matrix γ_{mn}^{-1} . If the bispectrum has no angular dependence then the inner product can be computed using the simplified expression (27), which yields

$$\beta_n^{Q, \text{theory}} = \frac{1}{8\pi^4} \sum_m \gamma_{nm}^{-1} \int_{\mathcal{V}} dk_1 dk_2 dk_3 \sqrt{k_1 k_2 k_3} B_e^{\text{theory}}(k_1, k_2, k_3) Q_m(k_1, k_2, k_3), \quad (31)$$

where we have used the quantity B_e defined in (16). The $\beta_n^{R, \text{theory}}$ may be obtained by the transformation $\beta_n^R = \sum_m \lambda_{mn} \beta_m^Q$. The appearance of the phase bispectrum in (31) is a consequence of our choice of weight w .

Equation (30) would continue to apply were we to change the definition of the ‘inner product’ $\langle\langle \cdot | \cdot \rangle\rangle$, and an analogue of (31) would continue to give the individual $\beta_n^{Q, \text{theory}}$. Our choice of signal-to-noise weighting in $\langle\langle \cdot | \cdot \rangle\rangle$ is important only for construction of the R modes and the covariance inherited by the $\beta_n^{R, \text{theory}}$.

3.4.3 Numerical evaluation

In practice, equation (31) requires evaluation of a 3-dimensional integral over the region \mathcal{V} . To implement it, we compute wB on a 200^3 cubic lattice in (k_1, k_2, k_3) and estimate the integral by volume-weighted cubature over this lattice. Some work is required to account for irregular boundary orientations; we give these details in Appendix A2.

3.4.4 Covariance

Finally we compute the covariance of estimators for the β_n^R coefficients under the assumption of Gaussian covariance for the bispectrum estimator $\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)V^{-1}\mathbf{1}_{\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3,0}$. Using equation (26), and (30) with R exchanged for Q , we obtain

$$\begin{aligned} \langle\langle \beta_m^R \beta_n^R \rangle\rangle &= (2\pi)^3 \delta(\mathbf{0}) \frac{6}{V^2} \frac{(8\pi^4)^2}{(k_{\max} - k_{\min})^6} \int \frac{d^3 k_1 d^3 k_2 d^3 k_3}{(2\pi)^9} (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \frac{R_m(k_1, k_2, k_3) R_n(k_1, k_2, k_3)}{k_1 k_2 k_3}, \\ &= \frac{6}{V} \frac{(8\pi^4)^2}{(k_{\max} - k_{\min})^6} \langle\langle R_m | R_n \rangle\rangle. \end{aligned} \quad (32)$$

The weighting for each Fourier configuration matches the signal to noise, making this correlator diagonal as a consequence of our construction of the R basis. Therefore we conclude

$$\text{Cov}_G(\beta_m^R, \beta_n^R) = \frac{6}{V} \frac{8\pi^4}{(k_{\max} - k_{\min})^3} \mathbf{1}_{mn}. \quad (33)$$

As for the line correlation function, it is independent of redshift and cosmological parameters. If we were to abandon the approximation of Gaussian covariance then (32) would no longer be proportional to exactly $\langle\langle R_m | R_n \rangle\rangle$. In this case the amplitude of the diagonal elements would be modified, and non-diagonal components would appear.

3.5 Galaxy bias

The discussion in Sections 3.2–3.4 was framed in terms of the dark matter overdensity δ , but this is not what is measured by surveys of a large-scale structure. Instead, they record the abundance of galaxies or some other population of tracers whose density responds to the dark matter density but need not match it.

On large scales the relation between the galaxy (δ_g) and dark matter (δ) density fields is well-described by the linear model $\delta_g = b_1 \delta$ (Kaiser 1984; Fry & Gaztanaga 1993). The *linear bias parameter* b_1 may be redshift-dependent, and varies between different populations of galaxies. On small scales the overdensities are larger, and both non-linear and non-local corrections become important. To obtain a satisfactory description, we must typically include terms at least quadratic (or higher) in δ (Fry & Gaztanaga 1993; Smith, Scoccimarro & Sheth 2007), together with terms involving the tidal gravitational field (Catelan, Porciani & Kamionkowski 2000; McDonald & Roy 2009; Chan, Scoccimarro & Sheth 2012; Baldauf et al. 2012).

In what follows, we assume the local Lagrangian bias model, in which the galaxy overdensity at early times is taken to be a local function of the dark matter overdensity. At later times the bias is determined by propagating this relationship along the dark matter flow. McDonald & Roy (2009) demonstrated that this implies the Eulerian galaxy overdensity at the time of observation can be written

$$\delta_g(\mathbf{x}) = b_1 \delta(\mathbf{x}) + \frac{1}{2} b_2 [\delta^2(\mathbf{x}) - \langle \delta^2(\mathbf{x}) \rangle] + \frac{1}{2} b_{s^2} [s^2(\mathbf{x}) - \langle s^2(\mathbf{x}) \rangle] + \dots, \quad (34)$$

where ‘...’ denotes terms of third order and higher that we have not written explicitly. The field $s^2(\mathbf{x}) = s^{ij}(\mathbf{x}) s_{ji}(\mathbf{x})$ is a contraction of the tidal tensor, defined by $s_{ij}(\mathbf{x}) \equiv [\partial_i \partial_j \nabla^{-2} - \frac{1}{3} \mathbf{1}_{ij}] \delta(\mathbf{x})$. Therefore, up to second order in δ , we require two additional redshift- and population-dependent bias parameters: the *quadratic bias* b_2 as well as the *non-local bias* b_{s^2} . In the local Lagrangian model the non-local bias satisfies $b_{s^2} = -4(b_1 - 1)/7$ (Chan et al. 2012; Baldauf et al. 2012), although in more general biasing prescriptions it could be allowed to vary independently.

3.5.1 Power spectrum

After translating to Fourier space, it follows that the tree-level galaxy power spectrum can be written

$$P_{\text{gal}}^{\text{tree}}(k) = b_1^2 P^{\text{tree}}(k). \quad (35)$$

To obtain a consistent result at one loop, we should include the unwritten third-order contributions in (34), which generate multiplicative renormalizations of the linear power spectrum in the same way as the ‘13’ terms of one-loop SPT. McDonald & Roy (2009) showed that these could be collected into a single new parameter that we denote $b_{3\text{nl}}$ to match Gil-Marín et al. (2015). Therefore

$$P_{\text{gal}}^{1\text{-loop}}(k) = b_1^2 P^{1\text{-loop}}(k) + 2b_1 b_2 P_{b_2}(k) + 2b_1 b_{s^2} P_{b_{s^2}}(k) + b_2^2 P_{b_2 b_2}(k) + 2b_2 b_{s^2} P_{b_2, b_{s^2}}(k) + b_{s^2}^2 P_{b_{s^2} b_{s^2}}(k) \\ + 2b_1 b_{3\text{nl}} \sigma_3^2(k) P^{\text{tree}}(k). \quad (36)$$

Saito et al. (2014) showed that in the local Lagrangian model $b_{3\text{nl}}$ satisfies $b_{3\text{nl}} = 32(b_1 - 1)/315$. Explicit expressions for all terms appearing in (36) were given by McDonald & Roy (2009). Note that contributions from the non-linear bias appear only in the one-loop power spectrum.

3.5.2 Bispectrum

In contrast to the power spectrum, the bispectrum receives corrections from non-linear bias terms even at tree level. Specifically,

$$B_{\text{gal}}^{\text{tree}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = b_1^3 B^{\text{tree}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + b_1^2 P^{\text{tree}}(k_1) P^{\text{tree}}(k_2) [b_2 + b_{s^2} S_2(\mathbf{k}_1, \mathbf{k}_2)] + \text{cyclic}, \quad (37)$$

where $S_2(\mathbf{k}_1, \mathbf{k}_2) \equiv (\mathbf{k}_1 \cdot \mathbf{k}_2)^2 / (k_1 k_2)^2 - 1/3$ is the kernel appearing in the Fourier transform of the contracted tidal field, $s^2(\mathbf{k}) = (2\pi)^{-3} \int d^3 q S_2(\mathbf{q}, \mathbf{k} - \mathbf{q}) \delta(\mathbf{q}) \delta(\mathbf{k} - \mathbf{q})$.

To obtain the galaxy bispectrum consistently at one loop, one should compute the dark matter overdensity to fourth order in perturbation theory and develop the bias expansion to the same order. This procedure has been adumbrated in the literature (Assassi et al. 2014) but not developed completely. Therefore to obtain an estimate of the one-loop bispectrum, we make the approximation

$$B_{\text{gal}}^{1\text{-loop}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = b_1^3 B^{1\text{-loop}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + b_1^2 P^{1\text{-loop}}(k_1) P^{1\text{-loop}}(k_2) [b_2 + b_{s^2} S_2(\mathbf{k}_1, \mathbf{k}_2)] + \text{cyclic}. \quad (38)$$

This is consistent with the prescriptions used by Gil-Marín et al. (2015) and Baldauf et al. (2016).

3.5.3 Application to bispectrum proxies

The outcome of this discussion is that, to predict the integrated bispectrum, line correlation function or modal bispectrum for the galaxy density field, we should make the replacements $P^{\text{theory}}(k) \rightarrow P_{\text{gal}}^{\text{theory}}(k)$ and $B^{\text{theory}}(k_1, k_2, k_3) \rightarrow B_{\text{gal}}^{\text{theory}}(k_1, k_2, k_3)$ where necessary in equations (12), (15) and (31).

To obtain theory predictions at tree level, we use equations (35) and (37), whereas to obtain predictions at one-loop we use equations (36) and (38). Finally, to evaluate predictions using the halo model we apply equations (36) and (37), but with $P^{1\text{-loop}} \rightarrow P^{\text{halo}}$ and $B^{\text{tree}} \rightarrow B^{\text{halo}}$ for the dark matter correlations.

Table 1. Fiducial values of the cosmological parameters, together with the step size $\Delta\theta$ used to vary each parameter in the simulations. We perform one simulation with offset $+\Delta\theta$ and one with increment $-\Delta\theta$, giving two offset simulations per parameter. With seven parameters and four realizations per model, this gives $4 + 2 \times 7 \times 4 = 60$ simulations in the suite. The bias parameters are assumed to be $b_1 = 1$ and $b_2 = 0$.

Parameter θ	Ω_m	Ω_b	w_0	w_a	σ_8	n_s	h
Fiducial value	0.25	0.040	-1.0	0.0	0.8	1.00	0.70
$\Delta\theta$	± 0.05	± 0.005	± 0.2	± 0.1	± 0.1	± 0.05	± 0.05

4 ESTIMATING BISPECTRUM PROXIES FROM N-BODY SIMULATIONS

In this section we briefly describe our N -body simulations and explain how they are used to estimate the Fourier bispectrum and its proxies ib , ℓ and β_n^Q .

4.1 Simulations

Our measurements are based on two sets of simulations: (1) 200 N -body simulations containing dark matter only, with a fixed choice of fiducial cosmological parameters, and (2) a total of 60 simulations constructed by varying one cosmological parameter at a time, with four realizations per model including the fiducial set. These simulations were performed on the ZBOX supercomputer at the University of Zurich and were described in Smith (2009) and Smith et al. (2014). Each set uses a comoving boxsize of $L = 1500 h^{-1}$ Mpc and contains $N = 750^3$ particles. Initial conditions for the particles were set at redshift $z = 49$ using 2LPT acting on a realization of a Gaussian random field (Crocce, Pueblas & Scoccimarro 2006) with transfer functions from CMBFAST (Seljak & Zaldarriaga 1996). The particles are evolved to $z = 0$ under the influence of gravity using the Gadget-2 code (Springel 2005), modified to allow a time-evolving equation of state for dark energy.

The fiducial cosmological parameters correspond to a flat Λ CDM model and are summarized in Table 1. Specifically, Ω_m and Ω_b are the matter and baryon density parameters; w_0 and w_a parametrize the equation of state for dark energy, viz. (Efstathiou 1999; Chevallier & Polarski 2001) $w(a) \equiv w_0 + (1 - a)w_a$; σ_8 is the amplitude of density fluctuations smoothed on a scale $8 h^{-1}$ Mpc; n_s is the spectral index of the primordial power spectrum; and h is the dimensionless Hubble parameter. We collectively write these as a vector θ_α with index α labelling the different parameters. To construct set (2), each parameter is offset by $+\Delta\theta_\alpha$ and $-\Delta\theta_\alpha$, with all other parameters held fixed. The step sizes $\Delta\theta_\alpha$ are listed in Table 1. To reduce noise when estimating parameter derivatives, we construct initial conditions for each of the four realizations using the same Gaussian random field as its fiducial partner. Since we vary over seven cosmological parameters, this gives a total of $4 + 2 \times 7 \times 4 = 60$ simulations in the suite.

4.2 Density field

To compute the overdensity field in each simulation, we use the cloud-in-cell assignment scheme to distribute particles over a regular Cartesian grid. We apply a fast Fourier transform and extract the discrete real-space density field by deconvolving the cloud-in-cell window function. The result is

$$\delta^{\text{disc}}(\mathbf{k}) = \frac{\delta^{\text{grid}}(\mathbf{k})}{W_{\text{CIC}}(\mathbf{k})}, \quad \text{where} \quad W_{\text{CIC}}(\mathbf{k}) = \prod_{i=1}^3 \left[\frac{\sin(\pi k_i / 2k_{\text{Ny}})}{\pi k_i / 2k_{\text{Ny}}} \right]^2. \quad (39)$$

The labels ‘disc’ and ‘grid’ label Fourier-space fields in the full volume V and on the cloud-in-cell grid, respectively. The Nyquist frequency $k_{\text{Ny}} = \pi N_{\text{grid}}/L$ is determined by the number of grid cells per dimension. For our numerical results, we use $N_{\text{grid}} = 512$.

4.3 Estimating the power spectrum

Given a realization of the δ -field within a simulation volume $V = L^3 = (2\pi)^3 \delta_D(\mathbf{0})$, a simple estimator for the power at wavevector \mathbf{k}_1 can be written $\hat{P}(\mathbf{k}_1, \mathbf{k}_2) = \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\mathbf{1}_{\mathbf{k}_1, -\mathbf{k}_2}/V$.⁴ Unfortunately this procedure is very noisy. An improved estimate can be obtained by summing over a set of modes satisfying the closure criterion $\sum_i \mathbf{k}_i = \mathbf{0}$ within a thin \mathbf{k} -shell. Since we are working in finite volume, the available modes are discretized in units of the fundamental frequency $k_f = 2\pi/L$, and therefore the thin-shell average should be written

$$\hat{P}(k) = \frac{1}{V_P(k)} \int d^3 q_1 d^3 q_2 \delta_D(\mathbf{q}_1 + \mathbf{q}_2) \hat{P}(\mathbf{q}_1, \mathbf{q}_2) \tilde{\Pi}_k(\mathbf{q}_1) \tilde{\Pi}_k(\mathbf{q}_2), \quad (40)$$

⁴ In the remainder of this paper we assume it is understood that we are dealing with the discrete density field whenever we refer to measured quantities, and drop the label ‘disc’.

where $\Delta k \geq k_f$ represents a bin width, and we have introduced the binning function $\tilde{\Pi}_k(\mathbf{q})$, which is defined to be unity if $|\mathbf{q}| \in [k - \Delta k/2, k + \Delta k/2]$ and zero otherwise. Finally, the quantity V_P represents the volume of the spherical shell accounting for discretization,

$$V_P(k) \equiv \int d^3 q_1 d^3 q_2 d^3 q_3 \delta_D(\mathbf{q}_1 + \mathbf{q}_2) \tilde{\Pi}_k(\mathbf{q}_1) \tilde{\Pi}_k(\mathbf{q}_2) = \int d^3 q \tilde{\Pi}_k^2(\mathbf{q}) = \int d^3 q \tilde{\Pi}_k(\mathbf{q}) = 4\pi k^2 \Delta k \left[1 + \frac{1}{12} \left(\frac{\Delta k}{k} \right)^2 \right]. \quad (41)$$

4.4 Estimating the bispectrum

In analogy with the power spectrum, an estimator for a single configuration of the Fourier bispectrum can be written $\hat{B}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \delta(\mathbf{k}_1) \delta(\mathbf{k}_2) \delta(\mathbf{k}_3) \mathbf{1}_{\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3, 0} / V$. (This expression was already used in Section 3.4 to obtain the Gaussian covariance for estimators of the β_n^R .) To obtain an acceptable signal to noise, we should again average over a set of configurations whose wavenumbers lie within suitable discretized \mathbf{k} -shells. After doing so, we obtain the estimator

$$\hat{B}(k_1, k_2, k_3) = \frac{1}{V_B(k_1, k_2, k_3)} \int d^3 q_1 d^3 q_2 d^3 q_3 \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \hat{B}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) \tilde{\Pi}_{k_1}(\mathbf{q}_1) \tilde{\Pi}_{k_2}(\mathbf{q}_2) \tilde{\Pi}_{k_3}(\mathbf{q}_3), \quad (42)$$

where the normalization V_B should now be evaluated using (Sefusatti et al. 2006; Joachimi, Shi & Schneider 2009)

$$V_B(k_1, k_2, k_3) \equiv \int d^3 q_1 d^3 q_2 d^3 q_3 \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \tilde{\Pi}_{k_1}(\mathbf{q}_1) \tilde{\Pi}_{k_2}(\mathbf{q}_2) \tilde{\Pi}_{k_3}(\mathbf{q}_3) \approx 8\pi^2 k_1 k_2 k_3 (\Delta k)^3. \quad (43)$$

Dividing by the square of the fundamental cell volume shows that the number of configurations scales as $N_{\text{triangles}}(k_1, k_2, k_3) = V_B(k_1, k_2, k_3) / k_f^6 \propto N_1 N_2 N_3$, where $N_i \equiv k_i / k_f$ is the length of the side k_i in units of the fundamental mode. Hence, if we scale the configuration by $k_i \rightarrow \lambda k_i$ then the number of available configurations scales as λ^3 .

Sefusatti (2005), Fergusson et al. (2012) and Scoccimarro (2015) observed that (42) could be implemented efficiently by rewriting the Dirac δ -function using its Fourier representation, $(2\pi)^3 \delta_D(\mathbf{q}) = \int d^3 x e^{i\mathbf{q} \cdot \mathbf{x}}$, and factorizing the dependence on the \mathbf{q}_i . This yields

$$\hat{B}(k_1, k_2, k_3) = \frac{k_f^3}{(2\pi)^6 V_B(k_1, k_2, k_3)} \int d^3 x \mathcal{D}_{k_1}(\mathbf{x}) \mathcal{D}_{k_2}(\mathbf{x}) \mathcal{D}_{k_3}(\mathbf{x}), \quad \text{where} \quad \mathcal{D}_k(\mathbf{x}) \equiv \int d^3 q e^{i\mathbf{x} \cdot \mathbf{q}} \delta(\mathbf{q}) \tilde{\Pi}_k(\mathbf{q}). \quad (44)$$

Similarly,

$$V_B(k_1, k_2, k_3) = \int \frac{d^3 x}{(2\pi)^3} \Pi_{k_1}(\mathbf{x}) \Pi_{k_2}(\mathbf{x}) \Pi_{k_3}(\mathbf{x}), \quad (45)$$

where $\Pi_k(\mathbf{x})$ is the inverse Fourier transform of $\tilde{\Pi}_k(\mathbf{q})$.

Equation (44) is numerically more efficient than a direct implementation of (42), because it requires only three Fourier transforms to compute \mathcal{D}_k for each wavenumber in the triplet $\{k_1, k_2, k_3\}$. Moreover, once each \mathcal{D}_k has been obtained, it can be re-used for any configuration that shares the same wavenumber. In spite of this improvement, however, it remains a formidable computational challenge to estimate all bispectrum configurations contained within a large volume V . Different strategies have been employed to make the calculation feasible. One option is to coarsely bin configurations with binning width equal to several times the fundamental mode. This drastically reduces the number of configurations to be measured. An alternative is to search only among a limited subset of configurations. This may be helpful if we wish to search for specific physical effects, but risks overlooking important signals if we are searching blindly. In either case the analysis is unlikely to be optimal because information is lost.

4.5 Estimating the integrated bispectrum

Our procedure to estimate the integrated bispectrum is based directly on its definition. We separate the total volume into N_s subvolumes, enumerated by the labels $i = 1, \dots, N_s$. We compute the mean overdensity $\hat{\delta}_i$ and power spectrum $\hat{P}(k)_i$ within each subvolume. Finally, we average the product $\hat{P}(k)_i \hat{\delta}_i$ over all subvolumes. Therefore,

$$\widehat{iB}(k) = \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{P}(k)_i \hat{\delta}_i. \quad (46)$$

The normalized integrated bispectrum can be obtained by rescaling,

$$\widehat{ib}(k) = \frac{\widehat{iB}(k)}{\hat{P}(k) \hat{\sigma}_L^2}, \quad (47)$$

where here $\hat{P}(k) = \sum_{i=1}^{N_s} \hat{P}(k)_i / N_s$ is the average subvolume power spectrum and $\hat{\sigma}_L^2 = \sum_{i=1}^{N_s} \hat{\delta}_i^2 / N_s$ is the average variance of the mean overdensity.

Table 2. Shell widths Δk used to average estimators for the power spectrum and bispectrum (where used), together with minimum and maximum modes k_{\min} , k_{\max} and the total number of bins or measurements N_{bin} .

	Δk [$h \text{ Mpc}^{-1}$]	k_{\min} [$h \text{ Mpc}^{-1}$]	k_{\max} [$h \text{ Mpc}^{-1}$]	N_{bin}
P	0.010	0.004	0.300	30
ib	0.010	0.021	0.306	29
ℓ	—	0.016	0.314	30
β	—	0.004	0.302	50
B	0.034	0.004	0.302	95

4.6 Estimating the line correlation function

A procedure to estimate the line correlation function was outlined by Eggemeier & Smith (2017). We evaluate

$$\hat{\ell}(r) = \left(\frac{r^3}{V}\right)^{3/2} \sum_{\substack{|\mathbf{k}_1|, |\mathbf{k}_2| \\ |\mathbf{k}_1 + \mathbf{k}_2| \leq 2\pi/r}} \bar{j}_0(|\mathbf{k}_1 - \mathbf{k}_2|r) \epsilon(\mathbf{k}_1) \epsilon(\mathbf{k}_2) \epsilon(-\mathbf{k}_1 - \mathbf{k}_2), \quad (48)$$

where $\bar{j}_0(|\mathbf{k}|r)$ denotes an average of $j_0(kr)$ taken over the volume of a fundamental k -space cell centred at \mathbf{k} . The sum scales as $\sim (2L/r)^6$, making its evaluation fast on large scales but challenging on small ones, where the sum includes the majority of Fourier modes. On scales below $\sim 105 h^{-1} \text{ Mpc}$, we find that the real space estimator described by Eggemeier & Smith (2017) becomes more efficient and therefore we use it within that regime. For scales accessible to both schemes, we verified that both estimators yield the same result.

4.7 Estimating the modal bispectrum

Equation (31) shows that an estimate of the modal coefficient β_m^Q requires evaluation of $\langle \langle w\hat{\mathcal{B}} | Q_n \rangle \rangle$, where $\hat{\mathcal{B}}$ is the bispectrum estimator defined in Section 4.4. Using equation (26), writing the δ -function using its Fourier representation, and factorizing the integral as described in Section 4.4, we find

$$\langle \langle w\hat{\mathcal{B}} | Q_n \rangle \rangle = \frac{1}{V} \int d^3x \mathcal{M}_{n_1}(\mathbf{x}) \mathcal{M}_{n_2}(\mathbf{x}) \mathcal{M}_{n_3}(\mathbf{x}), \quad \text{where} \quad \mathcal{M}_n(\mathbf{x}) \equiv \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k} \cdot \mathbf{x}} \frac{q_n(k)}{\sqrt{k \hat{P}(k)}} \delta(\mathbf{k}). \quad (49)$$

Here, $q_n(k)$ is a polynomial used in the construction of the modes Q_n ; see Appendix A1. Equation (49) shows that the computation can be reduced to a single 3-dimensional integral over the $\mathcal{M}_n(\mathbf{x})$, which are themselves weighted Fourier transforms of δ . Finally, β_m^Q can be estimated by contracting with the inverse inner product matrix γ_{mn}^{-1} defined in (28),

$$\hat{\beta}_m^Q = \sum_{n=0}^{n_{\max}-1} \langle \langle w\hat{\mathcal{B}} | Q_n \rangle \rangle \gamma_{nm}^{-1}. \quad (50)$$

To obtain the corresponding R -basis coefficients requires a further linear transformation

$$\hat{\beta}_n^R = \sum_m \lambda_{mn} \hat{\beta}_m^Q, \quad (51)$$

where λ_{mn} is the matrix defined above (29). As explained in Section 3.4, we generally perform numerical calculations in the Q basis in order to preserve the simplicity of (49), but present results in the R basis because their covariance properties make these coefficients simpler to interpret. In either basis, the measured coefficients can be used to reconstruct the bispectrum for any required Fourier configuration using equation (8).

Note that, because the matrix γ_{nm} can be tabulated, measuring a single modal coefficient has the same computational complexity as measuring a single configuration of the Fourier bispectrum.

4.8 Choice of bins

In Table 2 we summarize the parameters used in implementing estimators for each of these statistical quantities. The power spectrum and Fourier bispectrum are binned by averaging over shells of width Δk as explained in Sections 4.3–4.4. For the same reasons we also average the subvolume power spectra used to construct the integrated bispectrum. The line correlation function and modal coefficients do not involve averaging over shells, but instead are evaluated using equations (48) and (49), which are themselves aggregates over groups of configurations. For each statistic we report the minimum and maximum k -modes that contribute, and the total number of measurements or bins. Note that the bispectrum bin width corresponds to $\Delta k = 8 k_{\text{f}}$.

In what follows we will label the Fourier configurations for the bispectrum using the scheme of Gil-Marín et al. (2017). We assign the label (or ‘index’) zero to the equilateral configuration with $k_1 = k_2 = k_3 = k_{\min}$. The remaining configurations are ordered so that $k_1 \leq k_2 \leq k_3$ and $k_3 \leq k_1 + k_2$. Their labels are assigned by sequentially increasing k_3 , k_2 and k_1 (in this order) and incrementing the index for each valid triangle.

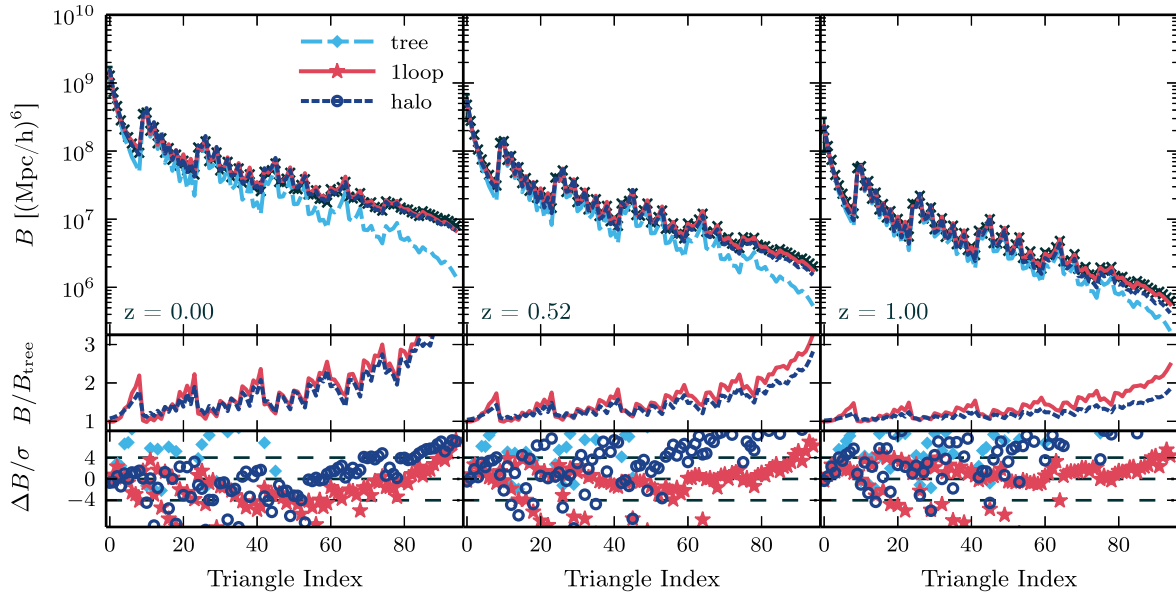


Figure 1. *Top row:* Measurements of the bispectrum as a function of configuration index (see the text), estimated from 200 N -body simulations at redshifts $z = 0, 0.52$ and 1 . We compare these measurements to the theoretical estimates of Section 3: the tree-level predictions are shown as dashed light-blue lines, the one-loop predictions are shown as solid red lines, and the halo model predictions are shown as short-dashed dark-blue lines. Black crosses mark the measured values. *Middle row:* One-loop and halo model predictions relative to the tree-level prediction. *Bottom row:* Differences between N -body measurements and theoretical predictions (i.e. $\Delta B = B^{\text{data}} - B^{\text{theory}}$), normalized to the corresponding 1σ standard deviation in the N -body value.

In our measurements of the integrated bispectrum, we split the simulation box into 125 subcubes, corresponding to a side of $300 h^{-1}$ Mpc. This increases k_{\min} by a factor of 5 compared to the full box. Finally, for the line correlation function we use a non-regular r -spacing, spanning the range from 10 to $200 h^{-1}$ Mpc. The first seven bins are separated by $2.5 h^{-1}$ Mpc, which doubles to $5 h^{-1}$ Mpc for the next 11 bins and to $10 h^{-1}$ Mpc for the remaining 12 bins.

5 COMPARISON OF THEORETICAL PREDICTIONS AND SIMULATIONS

In this section we present estimates of the typical values for each bispectrum proxy introduced in Section 2, and implemented using the formulae of Section 4. We derive these from the 200 simulations of our fiducial cosmology in set (1) – see Section 4.1 – at redshifts $z = 0$, $z = 0.52$ and $z = 1$. Also, using the simulation set (2), we determine how each proxy responds to changes in the cosmological parameters (Section 5.2). These measurements enable us to characterize the accuracy of the theoretical predictions for these typical values discussed in Section 3. Finally, in Section 5.3 we discuss measurements of the covariances and cross-covariances for each pair of proxies.

5.1 Mean values in the fiducial cosmology

5.1.1 Comparison of measurements and theoretical predictions

In Figs 1–4 we show measurements of each proxy for all three redshifts, averaged over the 200 different realizations. We do not explicitly display our power spectrum measurements, which have been well-studied by previous authors (e.g. Makino, Sasaki & Suto 1992; Lokas et al. 1996; Scoccimarro & Frieman 1996; Scoccimarro et al. 1998, 2001; Peacock & Smith 2000; Seljak 2000; Scoccimarro & Sheth 2002; Smith et al. 2003; Mead et al. 2015). In each figure, the top row contrasts our N -body measurements with the tree-level, one-loop and halo model predictions. The middle row displays the one-loop and halo model predictions relative to the tree-level prediction, and the bottom row shows the difference between the N -body measurements and the theoretical prediction in units of the standard deviation of the N -body estimate.

Fourier bispectrum. We find that both of the SPT predictions are more accurate at large scales and high redshifts. The halo model prediction is a better match at low redshift. The differences between each theoretical estimate and the typical values measured from simulation are broadly consistent with previous analyses; see Scoccimarro et al. (1998, 2001), Schmittfull et al. (2013), and Lazanu et al. (2016).

Integrated bispectrum. We give values for the normalized integrated bispectrum in Fig. 2. Except for a few k -bins the error bars are too large to show any preference for a particular theoretical model. In contrast to Figs 1, 3 and 4, the bottom row shows that tree-level SPT is a good match to the measured ib at all three redshifts. Conversely, the halo model prediction is a better match at high redshift. Our theoretical

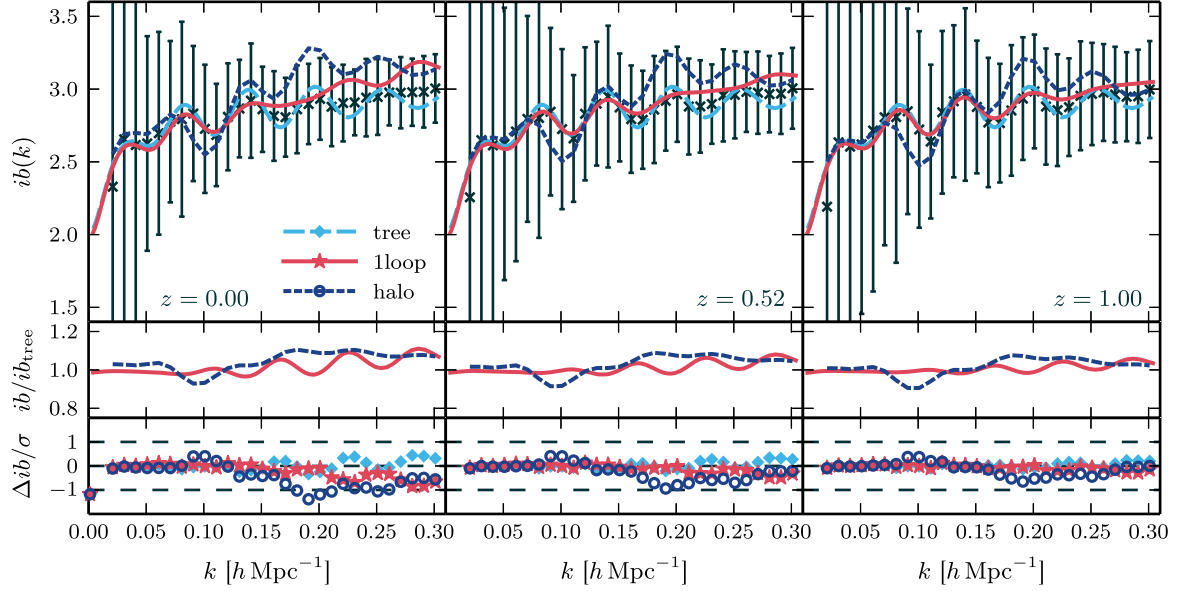


Figure 2. Same configuration as Fig. 1, showing values for the normalized integrated bispectrum. Error bars show the 1σ interval.

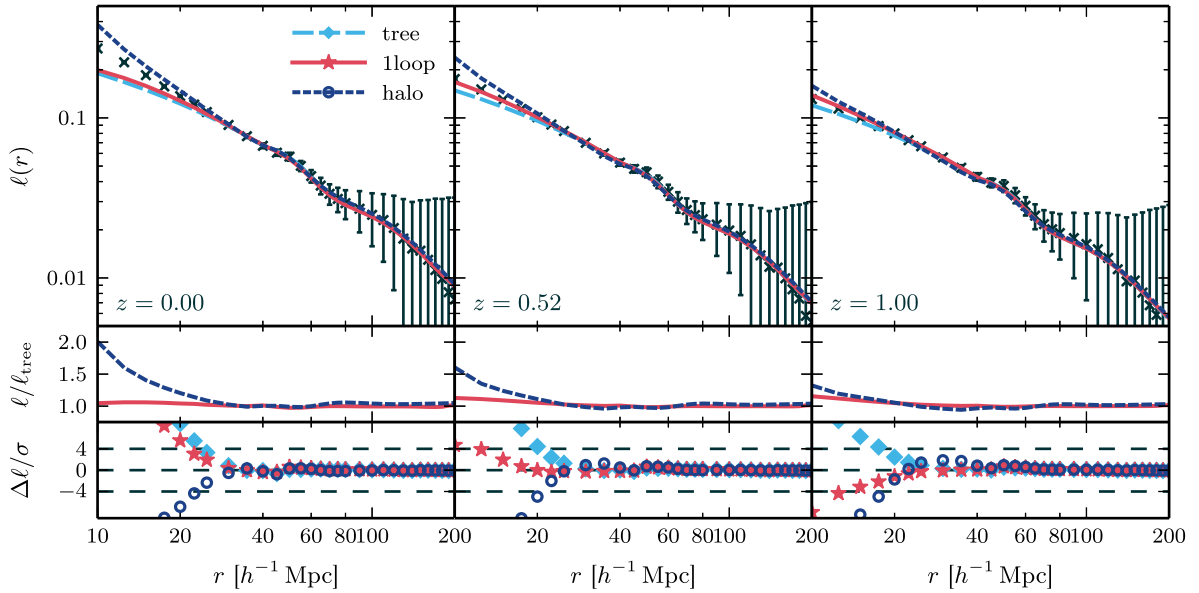


Figure 3. Same configuration as Fig. 1, showing values for the line correlation function at scale r . Error bars show the 1σ interval.

predictions are consistent with those reported by Chiang et al. (2014), but our measured values have larger error bars because we work with a smaller simulation volume.

Line correlation function. We present our measurements of the line correlation function in Fig. 3. The one-loop and halo-model predictions appearing here are new, and have not previously been studied. The most striking feature is the discrepancy between the halo model and SPT-based predictions in the smallest r -bins. This is consistent with the analyses of Wolstenhulme et al. (2015) and Eggemeier & Smith (2017), which both found differences between the tree-level prediction and values measured from simulation on scales with $r \lesssim 30 h^{-1} \text{ Mpc}$. The agreement is good for larger r .

Modal bispectrum. Finally, in Fig. 4 we plot the Fourier bispectrum reconstructed from (7) using our measurements of the β_n^Q coefficients. This is easier to interpret than the β -values themselves. The scatter between predicted and measured values (most clearly visible in the bottom row) is similar to the scatter for the directly measured Fourier bispectrum (Fig. 1), and indicates that differences between the reconstructed and directly-measured values are small. We give a more detailed analysis of the accuracy of the modal bispectrum in Section 5.1.2.

Theory error. The bottom panels of Figs 1–4 show that our theoretical predictions are accurate within a restricted range of scales. Outside this range it becomes progressively more difficult to model the observables. This mis-modelling should be regarded as an additional source

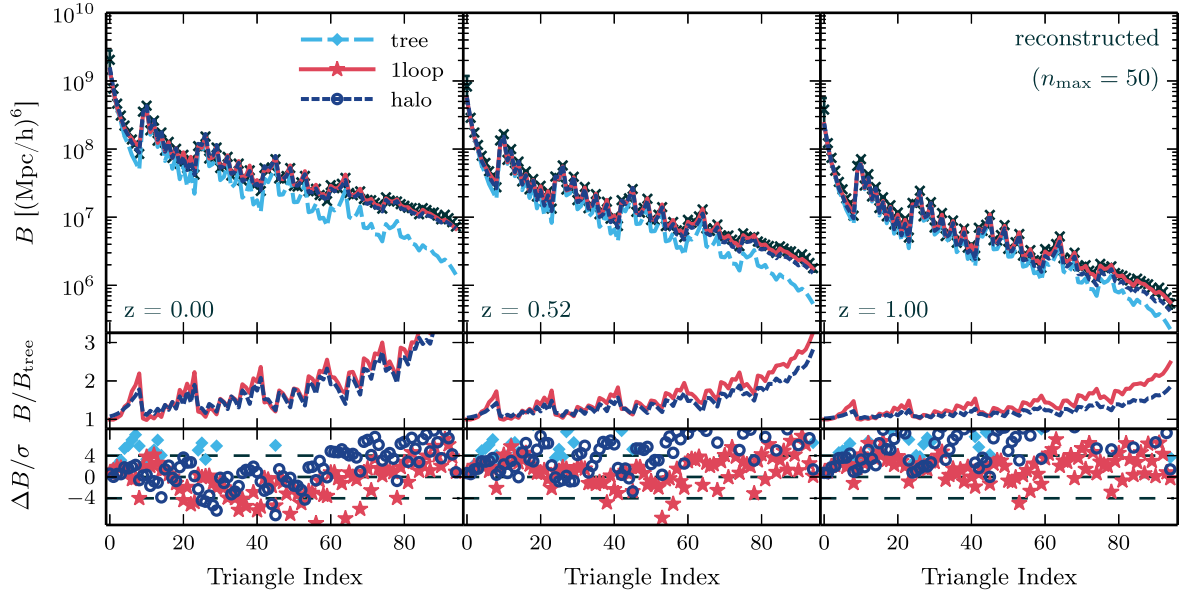


Figure 4. Same configuration as Fig. 1, showing values for the Fourier bispectrum reconstructed from the modal coefficients β_n^Q using equation (7). In the bottom row we plot differences computed using $\Delta B = B_{\text{modal}} - B^{\text{theory}}$.

of systematic error – a *theory error* – when forecasting constraints, or analysing data, using any of these theoretical models. In particle phenomenology, such theory errors are routinely estimated when performing fits to data, but their use in cosmology is less common. In this paper we construct Fisher forecasts for parameter error bars using both SPT-based models and the halo model. Comparison of these error bars enables us to estimate the impact of theoretical uncertainties on future constraints that incorporate 3-point statistics (see Section 7.4).

An alternative prescription for estimating theory errors was used by Baldauf et al. (2016) and Welling, van der Woude & Pajer (2016). In their approach the theoretical uncertainty in one-loop SPT is estimated from the next-order term in the loop expansion. We find that this prescription gives noticeably larger estimates than the difference between one-loop SPT and the values we measure from simulations. Therefore, although Baldauf et al. (2012) and Welling et al. (2016) concluded that (for example) constraints on some types of primordial non-Gaussianity would be weakened significantly after accounting for theory errors, our numerical comparison suggests that the attainable error may degrade by less than what their analysis would suggest.

5.1.2 Accuracy of modal reconstruction

Comparison of Figs 1 and 4 demonstrates that the Fourier bispectrum reconstructed from our measurements of the β_n^Q accurately reproduces the correct amplitude and shape dependence. This information is embedded in the modal coefficients. For example, the zeroth basis mode $R_0 \propto Q_0$ is a constant and therefore $\beta_0^R \propto \beta_0^Q$ captures information about the mean amplitude of the Fourier bispectrum over all configurations – or, equivalently, the skewness of δ . The next few modes are slowly varying functions of configuration. Taken together, these low-order modes carry the principal amplitude information and for reasonably smooth bispectra we expect they exhibit the strongest dependence on background cosmological parameters. The higher modes capture more subtle detail. As with any basis decomposition, their inclusion increases the accuracy of the reconstruction.

To see this in detail, consider a reconstruction using only $n_{\text{max}} = 10$ modes. In Fig. 5 we plot the Fourier bispectrum reconstructed in this way (blue line) compared to the reconstruction using $n_{\text{max}} = 50$ described above (red line). Black crosses mark the measured data points. In the lower panel we plot the ratio between these measured values and the reconstructions. It should be highlighted that the Fourier bispectrum values are aggregations within cubes of side $\Delta k = 8k_f$; on large scales (corresponding approximately to small triangle index) one should expect that the bispectrum value might be systematically biased away from the true value. However, for large triangle indices it is clear that the modal estimator faithfully reconstructs the underlying bispectrum. The accuracy is good whether we use $n_{\text{max}} = 10$ or $n_{\text{max}} = 50$, but the scatter is smaller for $n_{\text{max}} = 50$. We conclude that, in this case, the first 10 modes are sufficient to capture the main behaviour of the Fourier bispectrum, but extra modes are helpful if we wish to reproduce the precise configuration dependence to within $\lesssim 10$ per cent accuracy.

5.2 Derivatives with respect to cosmological parameters

In the remainder of this paper our aim is to obtain Fisher forecasts of error bars for a parameter set θ_α , where the index α labels one of the cosmological parameters of Table 1. For this purpose the role of a theoretical model is to predict the derivatives of observables with respect to each parameter, and the accuracy of the forecast depends on the reliability of these predictions. In this section we study how well our

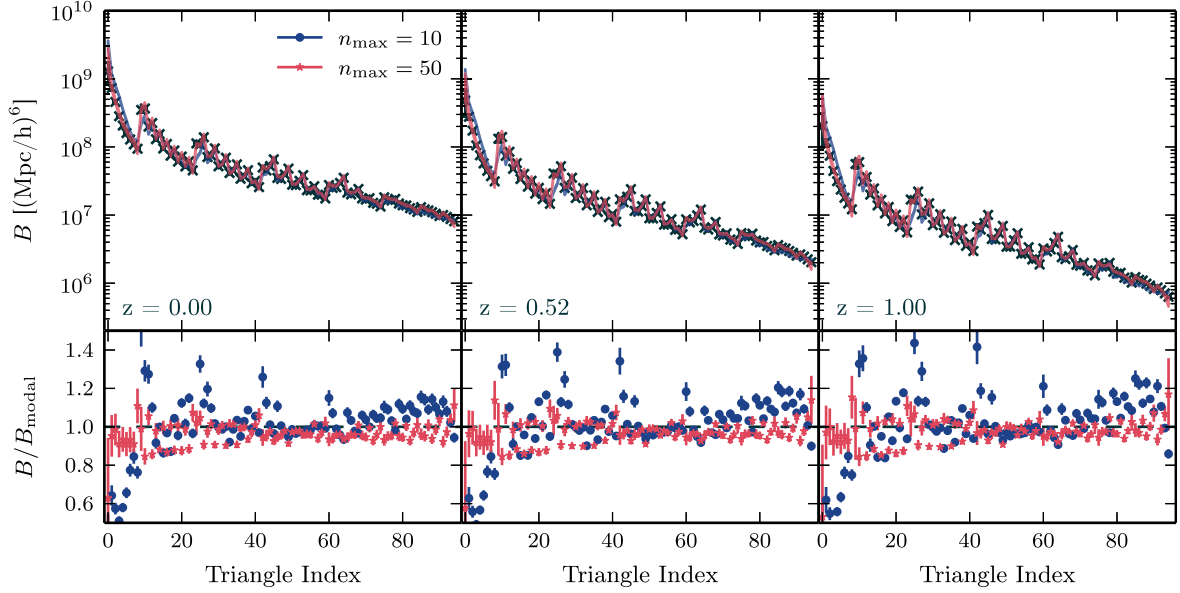


Figure 5. Modal bispectra reconstructed using 10 modes (blue) and 50 modes (red) at redshifts $z = 0, 0.52$ and 1 . The lower panels show the ratio of the measured normal bispectrum and modal bispectrum.

three theoretical models reproduce the derivatives estimated from our simulation suite. We compute the derivative of some estimator \hat{X} at wavenumber k with respect to a parameter θ_α by the rule

$$\frac{d\hat{X}(k|\theta)}{d\theta_\alpha} = \hat{X}(k|\theta) \frac{d \ln \hat{X}(k|\theta)}{d\theta_\alpha}, \quad (52)$$

where $\hat{X}(k|\theta)$ is the average over the 200 fiducial simulations of set (1) (described in Section 4.1) for $X \in \{P, B, \beta, ib, \ell\}$, and the logarithmic derivative with respect to θ_α is computed using

$$\frac{d \ln \hat{X}(k|\theta)}{d\theta_\alpha} = \frac{1}{4} \sum_{i=1}^4 \frac{\hat{X}^{(i)}(k|\theta + \Delta\theta_\alpha) - \hat{X}^{(i)}(k|\theta - \Delta\theta_\alpha)}{2\Delta\theta_\alpha \hat{X}^{(i)}(k|\theta)}. \quad (53)$$

The sum is over the four realizations used in a simulation set (2), and the derivative is constructed using the $+\Delta\theta_\alpha$ and $-\Delta\theta_\alpha$ offset simulations described in Section 4.1. The advantage of the logarithmic derivative is that both realizations in the numerator on the right-hand side of (53) share initial conditions with their fiducial partner in the denominator. Therefore, division by the fiducial estimate $\hat{X}^{(i)}(k|\theta)$ minimizes dependence on the specific realization.⁵

In Fig. 6 we plot the derivatives of each observable with respect to the cosmological parameters at $z = 0.52$. Our forecasts use three redshift bins, but their behaviour is similar to the $z = 0.52$ bin and the statements made below can be taken to apply to all three redshifts. We do not include the power spectrum, for which the derivatives appeared in Smith et al. (2014).

5.2.1 Integrated bispectrum

The derivatives of the integrated bispectrum are shown in the first column of Fig. 6. The error bars on the measured values are too large to show a clear preference for any model – and they are generally so large that the measurement is not significantly different from zero. These results are consistent with those reported by Chiang (2015) for a range of values of Ω_m , σ_8 and n_s . We conclude that the integrated bispectrum is rather insensitive to the background cosmology and is therefore a comparatively poor tool to constrain it. While this means we must expect a Fisher forecast to predict weaker error bars for the parameters of Table 1, this insensitivity could be an advantage if the intention is to use the integrated bispectrum as a probe of other physics. For example, in addition to the background cosmology, we may wish to use the large-scale structure bispectrum to constrain the possibility of *primordial* 3-point configurations produced by inflation on squeezed configurations. Insensitivity to the background cosmology would reduce the likelihood of degeneracies in these measurements.

5.2.2 Line correlation function

The second column of Fig. 6 shows the derivatives of the line correlation function. As for the typical values discussed above, the values predicted by our theoretical models are significantly discrepant with the measured values in the smallest r bins. Also, the derivative

⁵ This strategy is less successful for the line correlation function. In this case the fiducial value could be very close to zero on some scales. In turn, this produces large errors in the logarithmic derivative. Therefore, for the line correlation function, we estimate the linear derivative $d\ell/d\theta_\alpha$ instead.

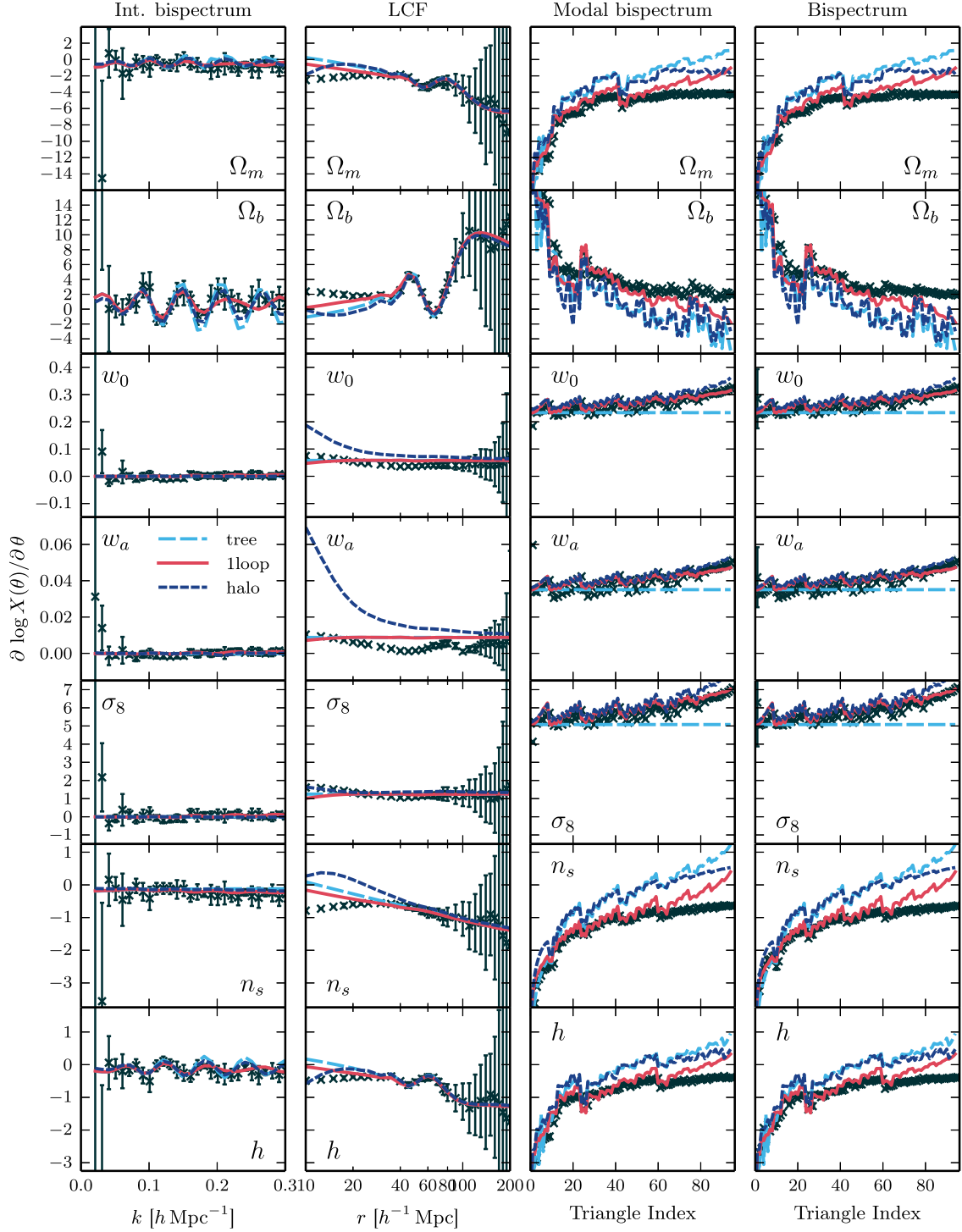


Figure 6. Derivatives of the Fourier bispectrum and its proxies with respect to the parameters at $z = 0$. The four columns show (from left to right) the derivatives of: ib , ℓ , B_{modal} (reconstructed from β) and B . As in Figs 1–4, measured values are in black, while light blue dashed, red solid and dark blue short-dashed lines are the tree-level, 1-loop and halo model predictions, respectively.

with respect to the dark energy parameter w_a is particularly discrepant for the halo model. One possible explanation is the construction of the halo model, with its fixed halo mass function and halo profile. Alternatively, it is possible that the halo model power spectrum and bispectrum that we use are subtly inconsistent in a way that produces inaccuracies in the line correlation function on small scales.

5.2.3 Modal bispectrum

To simplify comparison of the modal bispectrum with the Fourier bispectrum, Fig. 6 plots derivatives of the reconstructed bispectrum rather than derivatives of β_n^Q or β_n^R . Comparison of the final two columns shows that the cosmology dependence is accurately captured using $n_{\max} = 50$, either for theoretical predictions or for the measured values.

There is a significant spread in performance of the theoretical models, with tree-level SPT and the halo model generally offering the poorest match. For the derivatives with respect to Ω_m , Ω_b , n_s and h , these models give similar predictions. The probable reason is that, in the standard halo model, as alluded to above, the halo mass function and halo profile are fixed to the fiducial cosmology. Only the input power spectrum is taken to vary with the cosmological parameters, and since it matches the tree-level SPT prediction its derivatives will be equal. Therefore the halo-model derivatives will differ from those of tree-level SPT only via a (possibly scale-dependent) prefactor. More complex halo models with cosmology-dependent halo parametrizations have been studied (see e.g. Mead et al. 2016 for an application to dark energy models). However, determining which variation of the halo model captures the cosmological parameter dependence of the bispectrum most accurately is outside the scope of this paper. We simply note that if the halo model is to be used for analysis or forecasting of the Fourier bispectrum, its implementation should be chosen with care because its performance depends on these details.

5.3 Non-Gaussian covariance

The analytic, Gaussian covariance of each proxy is most accurate at high redshifts and on large scales, where the matter fluctuations are more nearly Gaussian and therefore more accurately described by the power spectrum alone. At low redshifts and on small scales, however, the Gaussian approximation fails due to non-linear evolution of matter fluctuations. This evolution generates additional contributions to the covariance through higher-order n -point correlations.

The simplest and most robust approach to obtain accurate non-Gaussian covariances has been to analyse large suites of N -body simulations. This method was used by Takahashi et al. (2009), Takahashi et al. (2011), Blot et al. (2016) and Klypin & Prada (2017) to study the non-Gaussian covariance of the power spectrum. Other authors have performed analogous studies for the bispectrum (Sefusatti et al. 2006; Chan & Blot 2016), the real-space partner of the integrated bispectrum (Chiang et al. 2015), and the line correlation function (Eggemeier & Smith 2017). In this section, we present our measurements of the non-Gaussian covariance for each proxy, estimated from our suite of simulations. We also discuss the cross-covariance between pairs of proxies.

In Sections 6 and 7, we quantify the impact of these complex non-diagonal covariances on estimates of signal to noise and Fisher forecasts.

5.3.1 Correlation matrices

We plot correlation matrices for the measurements $P + ib$, $P + \ell$, $P + \beta$ and $P + B$ in Fig. 7. We show measurements only at $z = 0$ where differences between the Gaussian and non-Gaussian covariances are largest.

The correlation coefficient r_{ij} between two data bins i and j is defined to satisfy $r_{ij} \equiv \hat{\mathbf{C}}_{ij} / \sqrt{\hat{\mathbf{C}}_{ii} \hat{\mathbf{C}}_{jj}}$, where $\hat{\mathbf{C}}$ is the covariance matrix estimated from the simulation suite,

$$\hat{\mathbf{C}}_{ij} = \frac{1}{N_{\text{real}}} \sum_{n=1}^{N_{\text{real}}} \left[\hat{S}_i^{(n)} - \bar{\hat{S}}_i \right] \left[\hat{S}_j^{(n)} - \bar{\hat{S}}_j \right], \quad (54)$$

and $N_{\text{real}} = 200$ is the number of realizations. To measure an auto-covariance, the data vector \mathbf{S} contains all measurements of a single proxy, $\mathbf{S} = (X_{a,1}, \dots, X_{a,n})$ or to measure a cross-covariance it contains all measurements from a pair, $\mathbf{S} = (X_{a,1}, \dots, X_{a,n_1}, X_{b,1}, \dots, X_{b,n_2})$, where $X_a, X_b \in \{P, B, \beta, ib, \ell\}$. The correlation matrix measures the degree of coupling between different measurements. Its elements take values between -1 (where the bins are fully anti-correlated) and $+1$ (where the bins are fully correlated). A value of 0 corresponds to independent measurements. For comparison, the Gaussian covariance matrices for P, B, β and ib are diagonal, whereas for ℓ there are correlations between neighbouring bins with similar r because it is a real-space statistic and therefore includes contributions from many Fourier configurations. In the Gaussian approximation, the cross-covariance between P and any bispectrum proxy is zero.

5.3.2 Integrated bispectrum and line correlation function

Correlation measurements for the integrated bispectrum appear in the top-left panel of Fig. 7. The $ib(k)$ measurements show stronger auto-correlations than $P(k)$ as k increases, while the $P \times ib$ cross-correlation is relatively featureless. This indicates that the two data sets are nearly independent. Similarly, we find that the $P \times \ell$ cross-correlation is nearly featureless except where the smallest r bins and highest k bins show significant correlation. Relative to the Gaussian covariance matrix for ℓ , the r bins with $r \lesssim 50 h^{-1}$ Mpc are more strongly correlated due to non-linear growth.

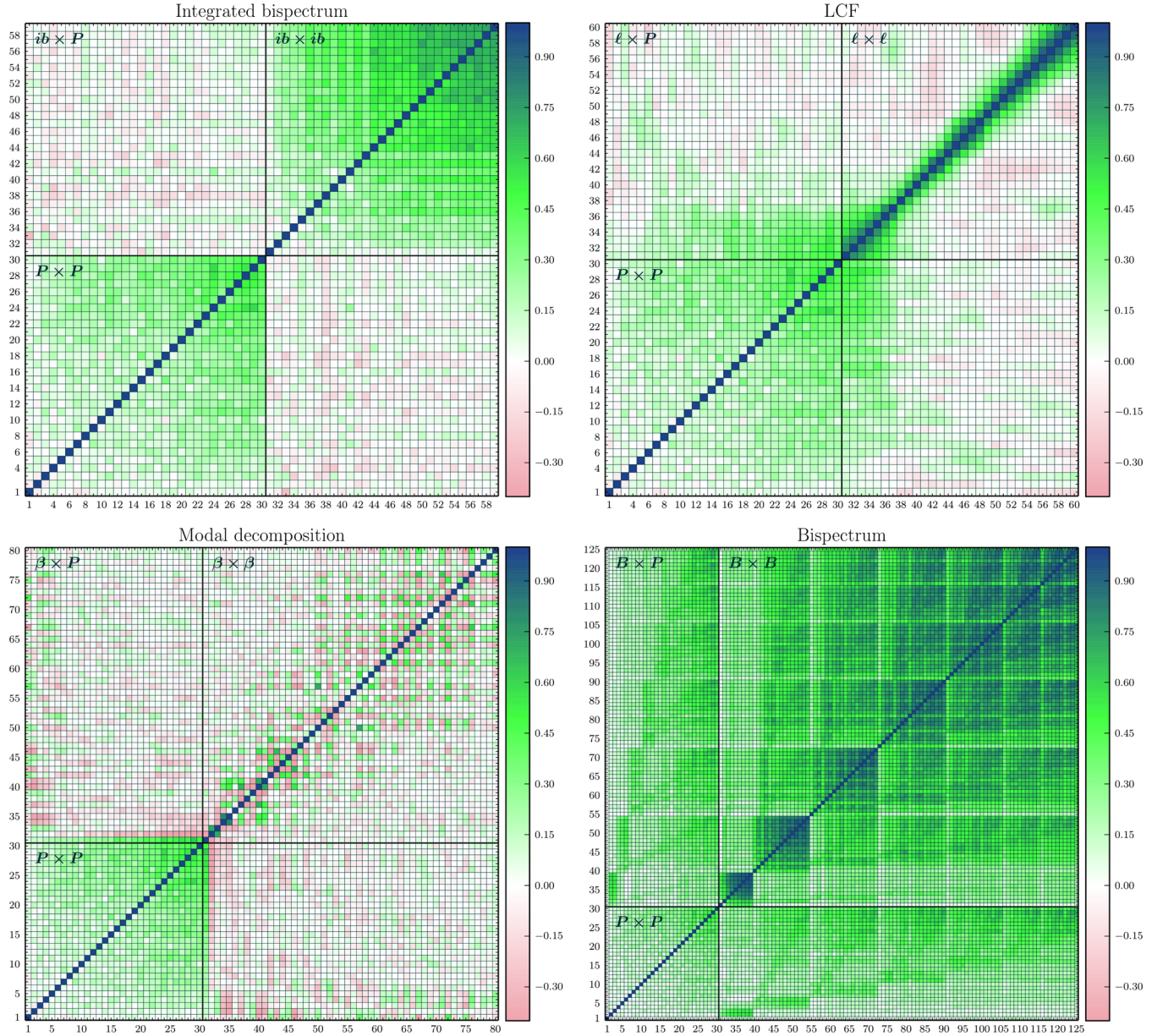


Figure 7. Correlation matrices for (clockwise from top left) $P + ib$, $P + \ell$, $P + B$ and $P + \beta$ at redshift $z = 0.0$. In each panel, the lower-left quadrant contains the power spectrum auto-correlation ($P \times P$), while the upper-right quadrant contains the auto-correlation of the corresponding 3-point correlation measure. The upper-left and lower-right quadrants contain the cross-covariance.

5.3.3 Modal bispectrum

In the lower-left panel of Fig. 7 we present measurements of the correlation coefficients for $P + \beta^R$. These have not previously been reported. As explained in Section 3.4 these measurements apply to the R basis, for which the covariance matrix is *constructed* to be diagonal in the Gaussian approximation. We find that only the first two modes are correlated with the majority of $P(k)$ bins. This is reasonable because the lowest modes probe the most scale-independent features of the phase bispectrum. The remainder show low-to-moderate correlation or anti-correlation due to non-linear effects.

5.3.4 Fourier bispectrum

For $P + B$ (bottom-right panel of Fig. 7) the correlation matrix has an approximate block structure due to the ordering of the 95 triangle configurations that we measure. The blocks correspond to groups of adjacent configurations with shared values of k_1 or k_2 . While the power spectrum $P(k)$ shows mild correlations between different bins at high k , the bispectrum exhibits much stronger correlations. There are also non-zero cross-correlations between power spectrum and bispectrum bins. The correlation between power spectrum and bispectrum tends to be higher when $P(k)$ and $B(k_1, k_2, k_3)$ have wavenumber bins that overlap. Similarly, the correlation between different bispectrum bins is

higher when the configurations share at least one wavenumber. However, even configurations that have no wavenumbers in common can be strongly correlated, with correlation coefficient as large as ~ 0.8 , due to non-linear growth.

5.3.5 Cross-covariances

Finally, we have computed the correlation matrices between the bispectrum and its proxies. These enable us to identify which bispectrum configurations contribute most to individual bins of ib , ℓ or β^R .

We find that B and ib are very weakly correlated, which we attribute to ib being dominated by more strongly squeezed triangles than any we include in the 95 measured configurations of B . The line correlation function is found to be correlated with a majority of bispectrum configurations when $r \lesssim 40 h^{-1} \text{ Mpc}$. This indicates that the line correlation function is sensitive to many different shapes of Fourier triangle. Finally, we find that the first two β^R modes are strongly correlated with the bispectrum over a large range of triangles, while the remainder are generally more correlated with triangles on the largest scales (i.e. lower triangle index). This structure is similar to the $P + \beta^R$ correlation matrix. We do not find particularly strong correlations for $\ell \times ib$, but $\ell \times \beta^R$ shows that the line correlation function at small r is highly correlated with the first two β^R modes. This is consistent with the observation that both are sensitive to a wide range of Fourier configurations.

6 CUMULATIVE SIGNAL TO NOISE OF THE BISPECTRUM PROXIES

Before discussing the constraining power of each proxy, we first compute the available signal to noise. This is an intermediate step that characterizes the significance with which measurements of each proxy can be extracted from a data set. Negligible signal to noise would normally imply poor prospects for parameter constraints. For example, Chan & Blot (2016) and Kayo, Takada & Jain (2013) studied the signal to noise as a proxy for the information content of the Fourier bispectrum in the context of a large-scale structure and weak lensing, respectively.

6.1 Numerical procedure

The cumulative signal-to-noise S/\mathcal{N} up to a maximum wavenumber k_{max} is defined by

$$\left(\frac{S}{\mathcal{N}}\right)^2 \equiv \sum_{k_i, k_j \leq k_{\text{max}}} S_i \mathbf{C}_{ij}^{-1} S_j, \quad (55)$$

where \mathbf{S} is the vector of typical values for either a single proxy or a combination of proxies, defined below equation (54). In this and subsequent sections we drop the use of a hat to denote an estimated value, and an overbar to denote a mean. The sum in (55) runs over all bins containing wavenumbers that satisfy the condition $k \leq k_{\text{max}}$. For the Fourier bispectrum a bin corresponds to a triplet of wavenumbers (k_1, k_2, k_3) , all of which are required to be smaller than k_{max} .

We use the non-Gaussian covariance matrix measured from simulations, described in Section 5.3, which we denote by \mathbf{C}_* . Its inverse \mathbf{C}_*^{-1} is not an unbiased estimator of \mathbf{C}^{-1} . A simple prescription to approximately correct for this bias is to rescale \mathbf{C}_*^{-1} by an Anderson–Hartlap factor (Anderson 2003; Hartlap, Simon & Schneider 2006), which yields

$$\mathbf{C}^{-1} \approx \frac{N_{\text{real}} - N_{\text{bin}} - 2}{N_{\text{real}} - 1} \mathbf{C}_*^{-1}, \quad (56)$$

where N_{real} is the number of realizations used to estimate the covariance matrix and N_{bin} is its dimensionality.⁶ Care should be taken when computing the numerical inverse \mathbf{C}_*^{-1} , especially for combinations of measurements with signals of widely disparate magnitude. To avoid issues associated with ill-conditioning, we first compute the correlation matrix $\mathbf{r}_{*,ij} = \mathbf{C}_{*,ij} / \sqrt{\mathbf{C}_{*,ii} \mathbf{C}_{*,jj}}$, whose entries lie between -1 and $+1$. We determine the inverse \mathbf{r}_{ij}^{-1} using a singular value decomposition and check that all singular values are above the noise. Finally, we compute the inverse covariance using

$$\mathbf{C}_{*,ij}^{-1} = \frac{\mathbf{r}_{*,ij}^{-1}}{\sqrt{\mathbf{C}_{*,ii} \mathbf{C}_{*,jj}}}. \quad (57)$$

6.2 Results

In Fig. 8 we plot the resulting signal-to-noise measurements for the integrated bispectrum, line correlation function, the quantity B_ϵ defined in (16) – and used in the construction of the line correlation function and the modal bispectrum – and the Fourier bispectrum. (The signal to noise from B_ϵ and the reconstructed modal bispectrum give almost identical results as, due to our choice of weighting function (see Eq. (24)), Eq. (8) implies $B_\epsilon(k_1, k_2, k_3) \approx \sum_n \beta_n^Q Q_n(k_1, k_2, k_3) / \sqrt{k_1 k_2 k_3}$.)

⁶ Although the Anderson–Hartlap prescription is simple to apply, it has been pointed out by Sellentin & Heavens (2016) that this rescaling simply broadens the Gaussian likelihood of the data. These authors argued that the distribution of the data is more accurately modelled by a t -distribution.

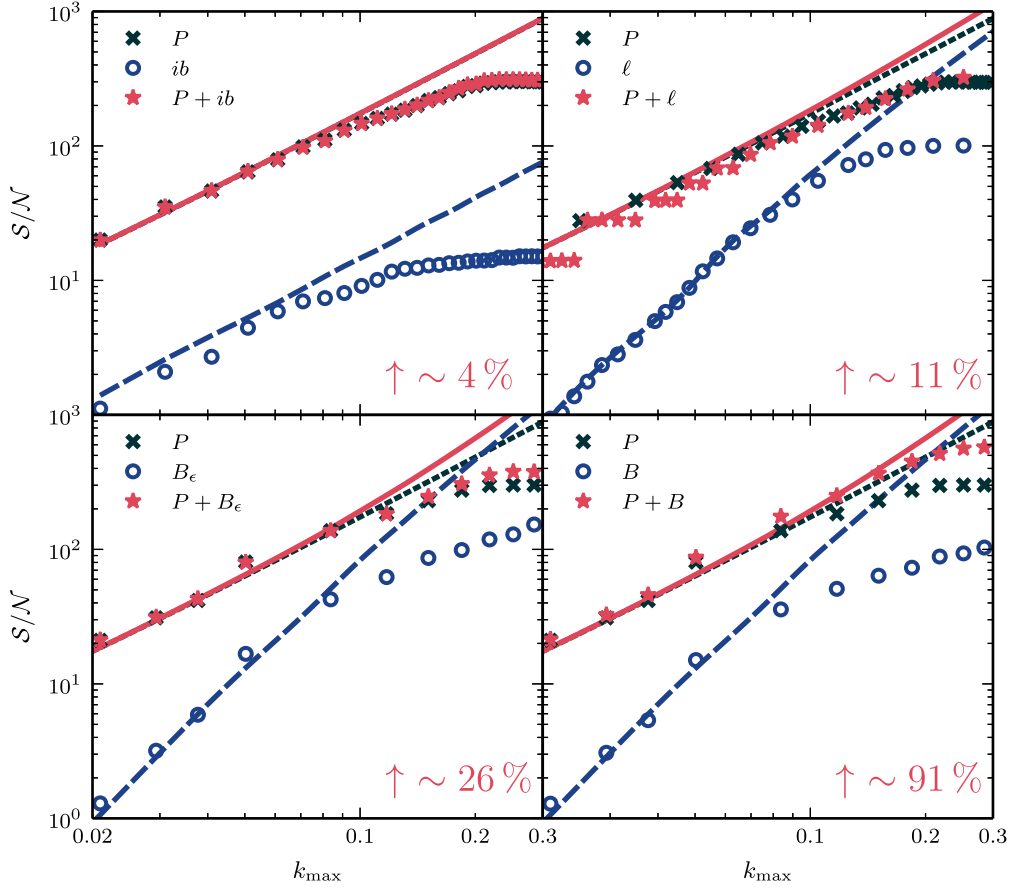


Figure 8. Cumulative signal to noise at redshift $z = 0$ as a function of the maximal mode k_{\max} for the measure X – equal to the integrated bispectrum, line correlation function, Fourier bispectrum or phase bispectrum (clockwise, starting from the top left panel). In each panel, blue circles refer to the measured signal to noise for X , while black crosses represent the signal to noise for the power spectrum. We plot the signal to noise for the combination $P + X$, including cross-covariance, as red stars. The blue, black and red lines give the theoretical prediction using the Gaussian approximation and tree-level SPT. The percentage quoted in the bottom right corner gives the increase in signal to noise relative to the power spectrum alone at $k_{\max} = 0.3 \, h \, \text{Mpc}^{-1}$.

Each panel of Fig. 8 shows the cumulative signal to noise of the Fourier bispectrum or a proxy (blue circles), together with the power spectrum (black crosses) and their combination, including the cross-covariance matrix (red stars). Our measurements of the integrated bispectrum and line correlation function carry forward the binning procedure used in Section 5. The step-like structure that occurs for $P + \ell$ is due to a mismatch of scales between the power spectrum and the bins of the line correlation function. The first four data points in the B and B_ϵ panels use a bin size $\Delta k = 2k_f$ in order to probe the low- k regime. The remainder derive from the measurements presented in Section 5 and use $\Delta k = 8k_f$. In each panel, for comparative purposes, we plot lines of matching colour to show the signal to noise computed using a Gaussian approximation to the covariance matrix and tree-level SPT to evaluate any correlation measures it contains.

6.3 Discussion

First, we note that the Gaussian approximation overpredicts the signal to noise for each proxy X and its combination $P + X$ with the power spectrum. This is consistent with the results reported by Chan & Blot (2016). The overprediction occurs because bins become coupled by non-linear evolution, and therefore do not provide independent information as the Gaussian approximation assumes. The effect can be quite severe: while the power spectrum signal to noise at $k_{\max} = 0.3 \, h \, \text{Mpc}^{-1}$ is overpredicted by a factor of 3, the impact on the Fourier bispectrum and its proxies is much larger. In these cases the overprediction ranges from a factor of ~ 5 or 8 for ib and ℓ up to more than an order of magnitude for the Fourier bispectrum. At smaller k_{\max} the overprediction is less, becoming significant for $k_{\max} \gtrsim 0.1 \, h \, \text{Mpc}^{-1}$.

The line correlation function, phase bispectrum and Fourier bispectrum *individually* contribute ~ 30 per cent of the signal to noise of $P(k)$ at $k_{\max} = 0.3 \, h \, \text{Mpc}^{-1}$, while the integrated bispectrum achieves only 5 per cent of the $P(k)$ signal to noise. For the Fourier bispectrum, this result is consistent with Chan & Blot (2016).

However, for estimating parameter constraints from the joint combination of P and B , or one of its proxies, the individual signal to noise contributed by one of these measurements is less important than whether it contains information that is not already present in the power spectrum. This is determined by the signal to noise of the combination $P + X$ compared to P alone. The different proxies show significant variation in the improvement from use of $P + X$, which we indicate as a percentage in the bottom-right corner of each panel. Although ℓ , B_ϵ

and B individually carry roughly the same signal to noise, the uplift in $P + X$ varies from ~ 11 per cent to ~ 91 per cent. Note that the signal to noise of $P + B$ receives a large improvement from the cross-covariance, which is in agreement with Chan & Blot (2016).

The discrepancy in uplift between B and B_ϵ is striking. If this discrepancy were to carry over to parameter constraints, it would imply that the Fourier bispectrum carries *significantly* more constraining power than B_ϵ , even though both statistics are equivalent in the approximation of Gaussian covariance. If true, this would be very surprising. We return to this question in Section 7.5 after we have obtained forecast parameter uncertainties for B and its proxies, which enable us to precisely quantify the constraining power of each statistic.

7 PARAMETER UNCERTAINTY FORECASTS

In this section we collect our major results, which are Fisher forecasts of the error bars achievable on the parameter set $\theta_\alpha = (\Omega_m, \Omega_b, w_0, w_a, \sigma_8, n_s, h)$ of Table 1, based on a fiducial flat Λ CDM cosmology. We perform these forecasts with and without inclusion of the bias parameters (b_1, b_2) .

In Section 7.1 we summarize our implementation of the Fisher forecasting method, and in Section 7.2 we present and compare the forecasts from each proxy. By comparing forecasts with and without non-Gaussian covariances, and using different theoretical models to describe the dark matter density, we are able to characterize their influence on the final parameter constraints. These discussions appear in Sections 7.3 and 7.4, respectively. Finally, we return to the discussion of Section 6 and examine to what extent the signal to noise provides a reliable metric by which to estimate improvements in parameter constraints (Section 7.5).

7.1 Forecasting method

The Fisher formalism can be used to forecast the precision with which cosmological parameters could be measured in a future survey. Consider a data vector \mathbf{x} containing measurements of any combination of statistical quantities. The likelihood function $\mathcal{L}(\theta | \mathbf{x})$ is defined to be the probability of the data, given the parameters θ , so $\mathcal{L}(\theta | \mathbf{x}) = P(\mathbf{x} | \theta)$. Then the Fisher matrix $\mathbf{F}_{\alpha\beta}$ satisfies

$$\mathbf{F}_{\alpha\beta} \equiv - \left\langle \frac{\partial^2 \ln \mathcal{L}(\theta | \mathbf{x})}{\partial \theta_\alpha \partial \theta_\beta} \right\rangle. \quad (58)$$

The expected 1σ error on each parameter θ_α , marginalized over all other parameters, can be obtained from the diagonal elements of the inverse Fisher matrix using $\sigma^2(\theta_\alpha) = (\mathbf{F}^{-1})_{\alpha\alpha}$. To simplify the computation of $\mathbf{F}_{\alpha\beta}$, we make the assumption that the likelihood function is a multivariate Gaussian,

$$\mathcal{L} = \frac{1}{\sqrt{(2\pi)^n |\mathbf{C}|}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad (59)$$

where \mathbf{T} denotes a matrix transpose and $|\mathbf{C}| = \det \mathbf{C}$ is the determinant of \mathbf{C} . We have written the mean of the data vector as $\boldsymbol{\mu} = \langle \mathbf{x} \rangle$, and its covariance matrix is $\mathbf{C}_{ij} = \langle x_i x_j \rangle - \mu_i \mu_j$. With these assumptions, it can be shown that (Tegmark, Taylor & Heavens 1997)

$$\mathbf{F}_{\alpha\beta} = \frac{1}{2} \text{tr} \left[\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_\alpha} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_\beta} \right] + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_\alpha} \mathbf{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_\beta}. \quad (60)$$

The first term measures variation of the covariance matrix with respect to the parameters, which is often a smaller effect than the variation of the means represented by the second term. In the approximation that this first term may be neglected, the Fisher matrix can be computed in terms of the inverse covariance matrix for the fiducial model. Our procedure to obtain this matrix from the simulation suite has already been described in Sections 5 and 6.

7.1.1 Survey configuration

The Fisher formalism depends explicitly on details of the survey under discussion, through both the specification of the data vector \mathbf{x} – such as how many redshift bins are used and which Fourier configurations are included – and the properties of the covariance matrix \mathbf{C} . In the following we adopt the parameters of an idealized survey of a large-scale structure consisting of three independent redshift slices at $z = 0$, $z = 0.52$ and $z = 1$. Each slice has volume $V = 3.375 h^{-3} \text{ Gpc}^3$ and a mode cut-off at $k_{\text{max}} = 0.3 h \text{ Mpc}^{-1}$. The total Fisher matrix can be written as a sum of the Fisher matrix in each slice,

$$\mathbf{F}_{\alpha\beta}^{\text{LSS}} = \mathbf{F}_{\alpha\beta}(z = 0) + \mathbf{F}_{\alpha\beta}(z = 0.52) + \mathbf{F}_{\alpha\beta}(z = 1). \quad (61)$$

We assume that in each redshift bin, the number density of galaxies is sufficiently high that the effect of shot noise is small. We do not include redshift-space distortions, the effect of complex survey geometry or the influence of super-survey modes. In general, all of these effects will be significant for a realistic survey and cannot be neglected. However, in this paper our intention is to address the question of whether the proxies described in Section 2 can be competitive with measurements of the Fourier bispectrum *in principle*. Survey-specific effects will generally reduce the number of configurations that can be measured, or increase the noise on those for which measurements are possible. This will typically weaken the performance of the proxies, meaning that their neglect gives us an estimate of the best-case scenario. While we do not anticipate that astrophysical or observational systematics will affect any one proxy more than the others, this is an interesting question to explore in future.

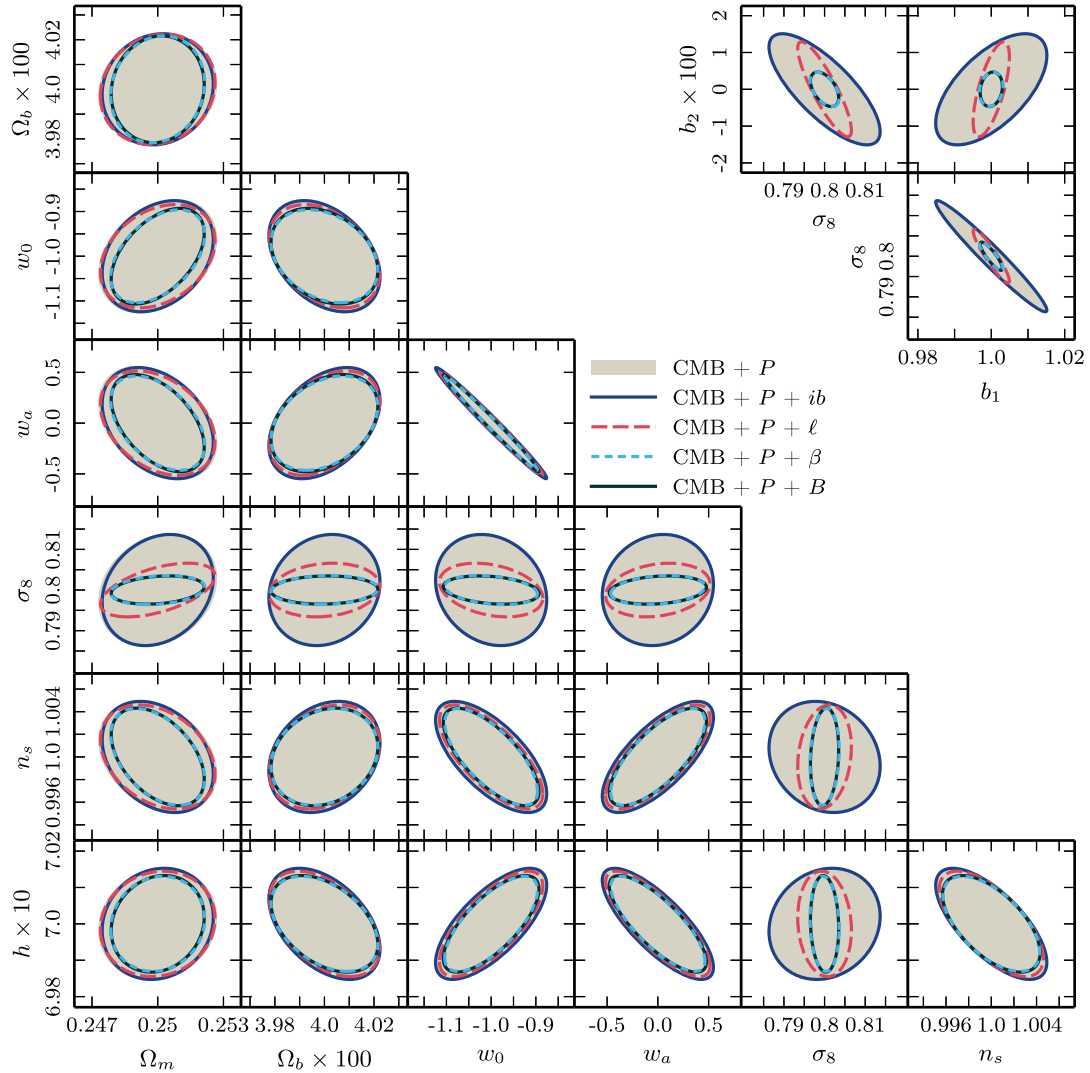


Figure 9. Comparison of marginalized 1σ likelihood contours forecast from a combination of the power spectrum and one of the following 3-point measures: integrated bispectrum (dark blue, solid), line correlation function (red, long-dashed), modal decomposition (light blue, short-dashed) and Fourier bispectrum (black, solid). The grey shaded regions show the error ellipses for the power spectrum alone. All forecasts include priors from a Planck-like CMB experiment and use a cut-off scale $k_{\text{max}} = 0.3 h \text{ Mpc}^{-1}$. The covariance matrices and parameter derivatives for the Fisher forecasts shown here are all derived from our simulation results in Section 5.

Each of the constraints we present includes a prior from the CMB power spectrum. We implement this prior by adding a fourth Fisher matrix,

$$\mathbf{F}^{\text{tot}} = \mathbf{F}^{\text{LSS}} + \mathbf{F}^{\text{CMB}}. \quad (62)$$

Details of the computation of \mathbf{F}^{CMB} for our choice of fiducial parameters were given by Smith et al. (2014).

7.2 Constraining power of the bispectrum and its proxies

In this section we present our forecasts. To minimize modelling errors, we construct the Fisher matrix for each proxy using quantities measured from simulation, except for derivatives with respect to the bias parameters that cannot be obtained in this way. For the Fourier bispectrum we compute these derivatives analytically by differentiating the one-loop power spectrum (36) and the tree-level bispectrum (37). Once the derivatives have been obtained, we replace occurrences of the dark matter power spectrum and bispectrum with their measured values. Our prescription for the proxies is similar, using the one-loop power spectrum to estimate derivatives of $P(k)$ and tree-level formulae together with the formulae of Section 3 to estimate derivatives of the proxy.

We plot the forecast 1σ confidence contours in Fig. 9. Each panel shows predicted joint constraints for a pair of parameters after marginalizing over all the others. The grey-shaded region marks the constraint predicted from measurements of the power spectrum only, except for inclusion of the CMB prior that we apply to all estimates. The solid dark-blue line marks the constraint predicted from $P + ib$; the long-dashed red line marks the constraint predicted from $P + \ell$; the short-dashed light-blue line marks the constraint predicted from $P + \beta$;

Table 3. Marginalized 1σ parameter uncertainties for the power spectrum and its combination with a 3-point correlation measure, including CMB priors. All quoted values are derived from the measured covariance matrices and parameter derivatives with $k_{\max} = 0.3 h \text{ Mpc}^{-1}$. The percentages in parentheses refer to the improvement over the P -only results.

	P	$P + B$			$P + \beta$ $n_{\max} = 50$		$P + \beta$ $n_{\max} = 10$		$P + \ell$		$P + ib$	
Ω_m	0.00179	0.00140	(22 per cent)	0.00141	(21 per cent)	0.00144	(19 per cent)	0.00172	(4 per cent)	0.00167	(7 per cent)	
Ω_b	0.00015	0.00014	(5 per cent)	0.00014	(5 per cent)	0.00014	(4 per cent)	0.00015	(2 per cent)	0.00015	(1 per cent)	
w_0	0.084	0.070	(16 per cent)	0.068	(19 per cent)	0.069	(17 per cent)	0.076	(9 per cent)	0.082	(2 per cent)	
w_a	0.370	0.315	(15 per cent)	0.306	(17 per cent)	0.310	(16 per cent)	0.338	(9 per cent)	0.360	(3 per cent)	
σ_8	0.0092	0.0023	(75 per cent)	0.0024	(74 per cent)	0.0025	(73 per cent)	0.0043	(53 per cent)	0.0090	(2 per cent)	
n_s	0.00327	0.00284	(13 per cent)	0.00281	(14 per cent)	0.00284	(13 per cent)	0.00303	(7 per cent)	0.00323	(1 per cent)	
h	0.00103	0.00087	(15 per cent)	0.00086	(16 per cent)	0.00087	(15 per cent)	0.00095	(7 per cent)	0.00101	(2 per cent)	
b_1	0.0103	0.0020	(81 per cent)	0.0021	(79 per cent)	0.0022	(79 per cent)	0.0032	(68 per cent)	0.0100	(3 per cent)	
b_2	0.0100	0.0031	(69 per cent)	0.0031	(69 per cent)	0.0031	(69 per cent)	0.0085	(15 per cent)	0.0100	(1 per cent)	

and the solid black line marks the constraint predicted from $P + B$. We summarize the marginalized 1σ error bars in Table 3. The value in parentheses following each uncertainty indicates the percentage improvement compared to the use of $P(k)$ alone.

7.2.1 Improvement from 3-point correlation data

First consider the joint constraints from $P + B$ (solid black lines in Fig. 9). These demonstrate that substantial improvements can be achieved compared to measurement of the power spectrum only. This is especially evident for σ_8 and the two bias parameters, for which the improvement is roughly 70–80 per cent; compare the second column of Table 3. This is perhaps unsurprising: the bispectrum constrains a different combination of σ_8 and b_1 than the power spectrum, and therefore assists in breaking their degeneracy (Fry 1994; Matarrese, Verde & Heavens 1997). Nevertheless, other parameters that do not participate in this degeneracy also experience improvements in the range 13–22 per cent, with the exception of Ω_b . This is already very well-measured by the CMB prior, and large-scale structure measurements can add little new information. These conclusions are similar to those reported by Sefusatti et al. (2006), who suggested that inclusion of Fourier bispectrum measurements could reduce uncertainties on Ω_m and σ_8 by a factor in the range 1.5–2.

Next, the forecast for the integrated bispectrum (solid dark-blue lines) shows that it offers negligible improvement, of order ~ 2 per cent, in comparison to P alone. This is consistent with the very small dependence on cosmological parameters discussed in Section 5.2, and the low signal-to-noise obtained in Section 6. On the other hand, the line correlation function offers comparable constraints to the Fourier bispectrum for σ_8 and b_1 , which receive improvements of 53 per cent and 68 per cent, respectively. Eggemeier & Smith (2017) demonstrated that this occurs because the line correlation function is nearly independent of b_1 and therefore probes a different direction in parameter space than P or B . Also, inclusion of ℓ measurements increases sensitivity to the dark energy parameters w_0 and w_a by ~ 9 per cent. These improvements are only marginally degraded compared to those from $P + B$, which are of order 15 per cent.

Finally, Fig. 9 demonstrates that the modal bispectrum with $n_{\max} = 50$ (short-dashed light-blue lines) is predicted to yield error bars nearly equivalent to the Fourier bispectrum with 95 triangles. Indeed the modal estimator for some parameters does slightly better than the bispectrum, which may lose some information due to the large bin width $\Delta k = 8k_f$. Nevertheless, there is no sign of the significant difference in constraining power between B and B_e – which is the quantity implicitly measured by β with our choice of basis – that was suggested by our analysis of signal to noise in Section 6. We return to this apparent discrepancy in Section 7.5 below. Just as important, the differences between the cases $n_{\max} = 10$ and $n_{\max} = 50$ are mostly negligible. Therefore, even with as few as $n_{\max} = 10$ modes, the modal decomposition retains nearly the full constraining power of the bispectrum. However, it should be remembered that Fig. 5 suggests the Fourier bispectrum reconstructed with so few modes will introduce more significant scatter. In a realistic analysis, these reconstruction errors could manifest themselves as a bias on the best-fitting cosmological parameters. Unfortunately we cannot account for this bias in our Fisher analysis, but it deserves further investigation.

7.2.2 Combination with other observables

The strong degeneracy between σ_8 and b_1 can be broken by other means. For example, it is possible to use weak lensing measurements that probe the matter power spectrum directly. Given that inclusion of 3-point correlation data yields the largest improvements for σ_8 and the bias, it is worthwhile considering what improvements should be expected were the bias to be fixed by other cosmological observations.

In a scenario of this kind the power spectrum constraints would not be weakened by marginalization over the bias parameters, and therefore inclusion of 3-point correlation data would no longer yield such a dramatic improvement for σ_8 . However, we still find encouraging improvements for many parameters. For example, inclusion of either Fourier or modal bispectrum measurements would decrease uncertainty on σ_8 by ~ 25 per cent and all other parameters except Ω_b by 10–15 per cent. Inclusion of ℓ measurements would decrease uncertainty on σ_8 by 20 per cent, on the dark energy parameters by ~ 10 per cent, and for all other parameters by $\lesssim 5$ per cent. We conclude that even in the extreme case that b_1 and b_2 can somehow be determined exactly, inclusion of 3-point correlation data still provides valuable additional information.

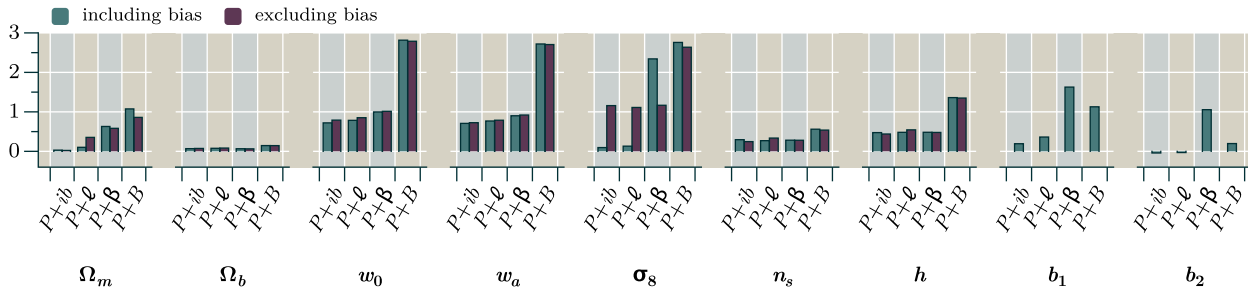


Figure 10. Increase in parameter uncertainties from non-Gaussian covariances, measured using $\sigma_{\text{NG}}/\sigma_{\text{G}} - 1$, where σ_{NG} (σ_{G}) is the predicted error bar using the non-Gaussian (Gaussian-like) covariance from simulations. Predictions that include (do not include) a marginalization over the bias parameters are in blue-green (purple).

These Fisher forecasts should be interpreted with some care. As explained above, we do not include a number of astrophysical and observational effects that complicate the analysis of realistic galaxy survey data. These include redshift uncertainties, redshift-space distortions, irregular survey geometries and shot noise. In particular, for the forecasts presented here the effective shot noise is set by the number density $\bar{n} = 0.125 h^3 \text{Mpc}^{-3}$ of particles in our simulation suite. This is substantially larger than the galaxy number densities that will be achieved by upcoming surveys. We return to this issue in Section 8.2, where we discuss how our predictions would be modified by a more realistic number density.

7.3 Effect of non-Gaussian covariance and cross-covariance

The non-Gaussian covariance measured in simulations differs from the Gaussian approximation in two ways: (1) it includes additional contributions to the variance of each bin from higher order correlations and (2) it adds or enhances coupling between different bins of a single proxy, and between bins of different proxies. These non-Gaussian corrections generally lead to weaker parameter constraints when compared to forecasts constructed using the Gaussian approximation, because this assumes that every bin contributes independent information. In this section, we compare the relative impact of non-Gaussian covariance for the different proxies by contrasting Fisher forecasts made with and without its inclusion. We give results for the combinations $P + ib$, $P + l$, $P + \beta$ and $P + B$ and each choice of theoretical model – tree-level SPT, 1-loop SPT or the halo model. For all forecasts in this work with (or without) non-Gaussian covariances, the non-Gaussian contributions to the covariance are always included (or ignored) in both P and the higher-order correlation statistic at the same time.

7.3.1 Increase in uncertainty from non-Gaussian contributions

Fig. 10 shows the relative increase $\sigma_{\text{NG}}/\sigma_{\text{G}} - 1$ in predicted uncertainty for each parameter when non-Gaussian contributions are included. To estimate σ_{G} , we use the expressions for Gaussian covariance given in Section 3, with each quantity replaced by its value measured from our simulations. For example, to construct the Gaussian covariance for ib , we use equation (14) with σ_L^2 replaced by its measured value. We could equally well have constructed similar estimates using one of the theoretical models to calculate such values, but the result is not very different. The discussion in this section would continue to apply if we were to reproduce Fig. 10 using estimates generated by any of these prescriptions.

The increase in uncertainty induced by inclusion of non-Gaussian effects depends on the measure of 3-point correlations used to generate constraints, the method used to estimate the Gaussian covariance matrix, and the parameter in question. In general we find that the Gaussian approximation underpredicts the uncertainty for the Fourier bispectrum more strongly than for its proxies. Note also that – although $P + \beta$ and $P + B$ yield nearly identical constraints when the non-Gaussian covariance is used, as described in Section 7.2 – the importance of the non-Gaussian covariance for these combinations is not the same. Since the quantity B_ϵ measured by β is not the same as B , neglecting cross-covariance with P (as the Gaussian covariance does) will leave out different information for $P + \beta$ compared to $P + B$.

Inclusion of non-Gaussian covariance impacts uncertainties for w_0 , w_a and σ_8 more significantly than the other parameters. This non-uniformity means that it is not obvious how inclusion of non-Gaussian covariance might impact constraints from 3-point correlations on further parameters not considered here. For instance, a number of authors have used Gaussian covariances to forecast future constraints on a primordial bispectrum generated by inflation; see Scoccimarro, Sefusatti & Zaldarriaga (2004), Sefusatti & Komatsu (2007), Sefusatti, Crocce & Desjacques (2012), Baldauf et al. (2016), Welling et al. (2016) and Tellarini et al. (2016). It is not yet clear how these forecasts will change when more realistic non-Gaussian covariances are used.

7.3.2 Inclusion of cross-covariance

In Fig. 11 we summarize the influence of cross-covariance between P and the 3-point measures by comparing constraints using the full non-Gaussian covariance to constraints where the cross-covariance has been set to zero. We find that inclusion of cross-covariances *reduces* the predicted uncertainties for nearly all parameters and choices of combination $P + X$, whether or not we marginalize over galaxy bias. In the few cases where inclusion of cross-covariance did not reduce the uncertainties (e.g. constraints on Ω_m from $P + B$ and $P + \beta$), the predicted

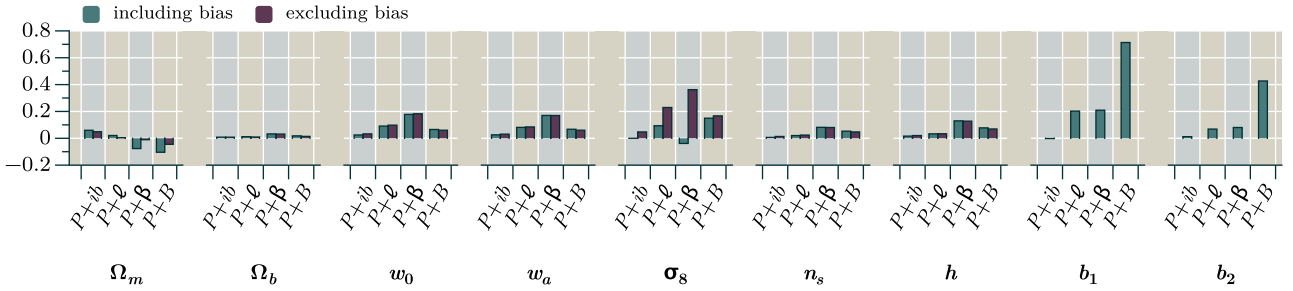


Figure 11. Improvement in parameter uncertainties from the inclusion of cross-covariance, measured using $\sigma_{\text{NG} - \text{no} - \text{CC}} / \sigma_{\text{NG} - \text{with} - \text{CC}} - 1$, where $\sigma_{\text{NG} - \text{with} - \text{CC}}$ is the error bar predicted using non-Gaussian covariance measured from simulations and $\sigma_{\text{NG} - \text{no} - \text{CC}}$ is the error bar predicted from the same covariance matrix, except with cross-covariances between P and each 3-point statistic set to zero. Predictions that include (do not include) a marginalization over the bias parameters are in blue-green (purple).

error bar is weakened by less than 12 per cent of the error bar without cross-covariance. Overall, we find that ignoring cross-covariances can overestimate uncertainties by up to ~ 40 per cent when we do not marginalize over the bias, and by 40–70 per cent for the special case of bispectrum constraints on the bias parameters themselves.

This reduction of uncertainties due to inclusion of cross-covariances may be surprising. While we have not explicitly identified the source of the improved constraining power, this is not a new feature of Fisher forecasts using non-Gaussian covariances. For example, a number of authors using cross-correlations between cluster counts, weak lensing power spectra and the weak lensing bispectrum have found that parameter constraints can improve when cross-covariances between strongly coupled measurements are included (Takada & Bridle 2007; Kayo et al. 2013; Sato & Nishimichi 2013). But it is also possible that our improvements are partly due to the galaxy biasing model we have chosen. A simulation of haloes, rather than dark matter alone, could be used to verify the effect when simultaneously constraining both cosmological parameters and galaxy bias.

The conclusion of this discussion is that an accurate estimate for the covariance matrix, including non-Gaussian contributions and off-diagonal terms, is important if we wish to obtain reliable constraints. Unfortunately, this is especially true for the Fourier bispectrum for which the Gaussian approximation most significantly underestimates the true parameter uncertainties. This implies that surveys aiming to generate constraints from inclusion of B measurements cannot evade the computational difficulties associated with estimating their covariance matrix.

To mitigate these difficulties, we could consider the use of $P + \beta$ rather than $P + B$. As we have seen in Section 7.2, these combinations yield nearly equivalent constraints using 95 Fourier configurations and 50 modal coefficients, respectively, and therefore the modal decomposition makes the information content of the bispectrum more accessible by reducing the size of the covariance matrix needed to obtain it. We consider the efficiency with which each proxy can compress the information carried by B in Section 8.1.

7.4 Theory dependence of the forecasts

In Section 7.2 we have presented our Fisher forecasts based on simulated data, and in Section 7.3 we have discussed the influence of non-Gaussian covariance and cross-covariances. These results enable us to assess the information content carried by the Fourier bispectrum and its proxies, but the question of how easily these statistics can be deployed remains open. In particular, we would like to know whether the use of simulated data is essential, or whether any of the models described in Section 3 are sufficient. In this section we study the dependence of our forecasts on the choice of theoretical model used to estimate the derivatives $\partial\mu/\partial\theta_\alpha$ in equation (60).

7.4.1 Match to forecast from simulations

First, we consider whether there is a model that provides a clear best-match to the forecast using simulated data. Fig. 12 compares the forecasts for each parameter using different prescriptions for the covariance matrix and for different choices of theoretical model, with marginalization over the bias included or excluded. The bar heights represent the reduction in the predicted uncertainty provided by a given combination, relative to the base model of power spectrum data only combined with a CMB prior. The results of Section 7.2 are labelled ‘sim’. Unfortunately, for each combination $P + X$ there is no single choice of theoretical model yielding forecasts that provide the best match to the ‘sim’ outcome for all parameters – with or without marginalization over bias.

For example, consider the combination $P + B$ in the first column of Fig. 12. This summarizes forecasts generated by including non-Gaussian covariance and marginalization over the bias. For σ_8 it is 1-loop SPT that gives the best match to the ‘sim’ result, but for the linear bias parameter b_1 the best match comes from tree-level SPT.

Alternatively, one could ask whether any one model provides uniformly conservative or uniformly optimistic forecasts. If so, that model could be used to estimate upper or lower limits on the uncertainty for any chosen parameter. But Fig. 12 demonstrates that there are no models with such properties. For example, focusing again on the first column, there is no single choice of theoretical model for $P + B$ that forecasts the largest or smallest improvement for all parameters.

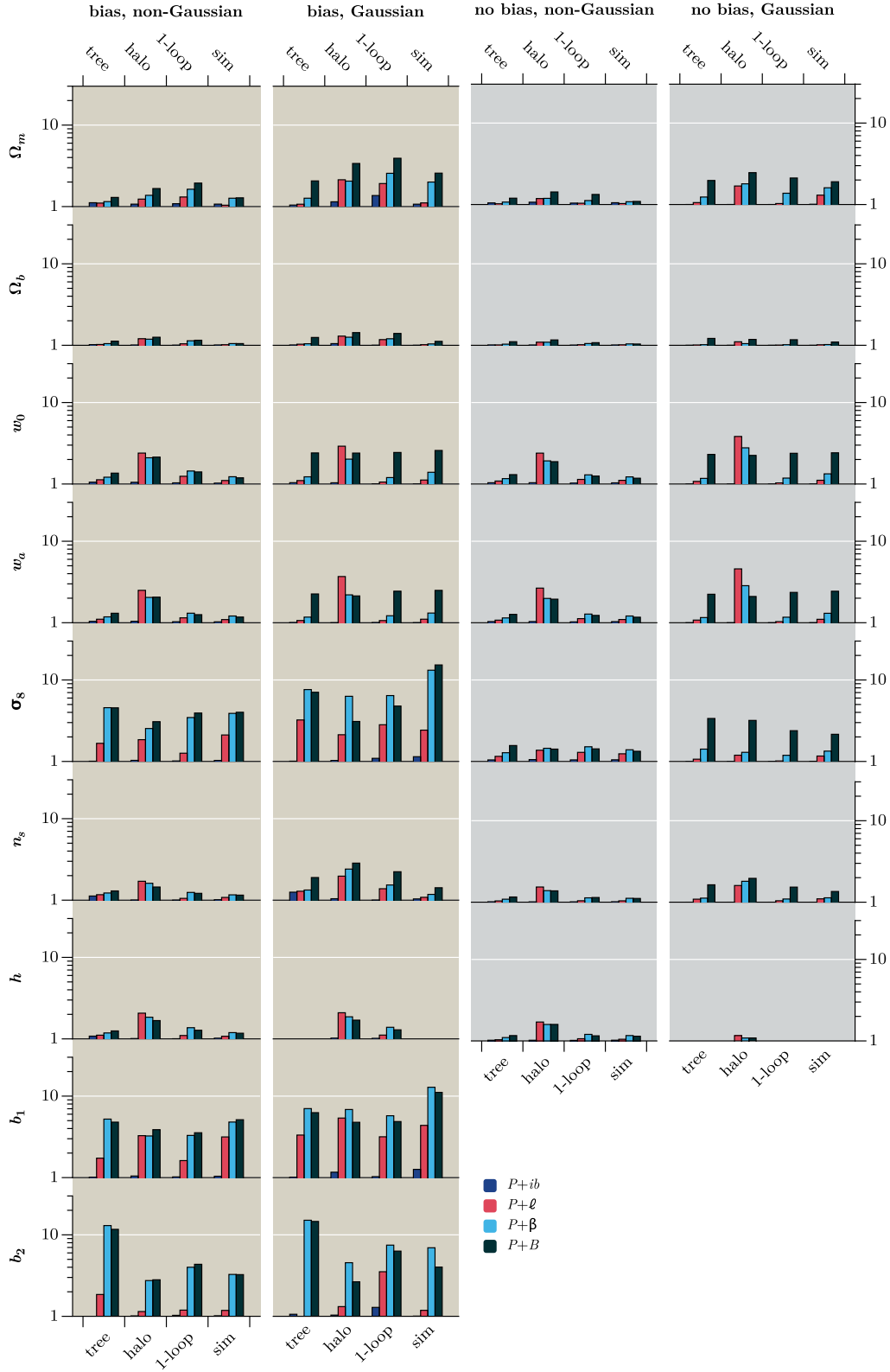


Figure 12. Improvements in parameter uncertainties from the addition of 3-point statistics are shown as bars with height $\sigma(P)/\sigma(P+X)$, the ratio of parameter errors from P only and the combination $P+X$. The labels at the top of each column indicate whether bias parameters are included or excluded, and whether Gaussian or non-Gaussian covariances are used. Each group of four bars corresponds to a different choice of theoretical model, and the colour of each bar indicates the $P+X$ combination. We note that, since the tree-level power spectrum does not depend on b_2 , for the two tree-level bar groups in the last row, the bar heights measure $\sigma_{\max}/\sigma(P+X)$, where σ_{\max} is the maximum error on b_2 among the four $\sigma(P+X)$ values.

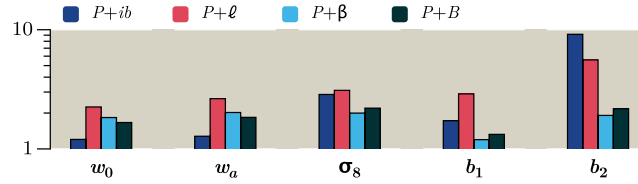


Figure 13. Sensitivity factors, defined as the ratio between the largest and smallest forecast parameter uncertainties among the three theoretical models, for each $P + X$ combination. The forecasts compared here include bias parameters and use non-Gaussian covariances.

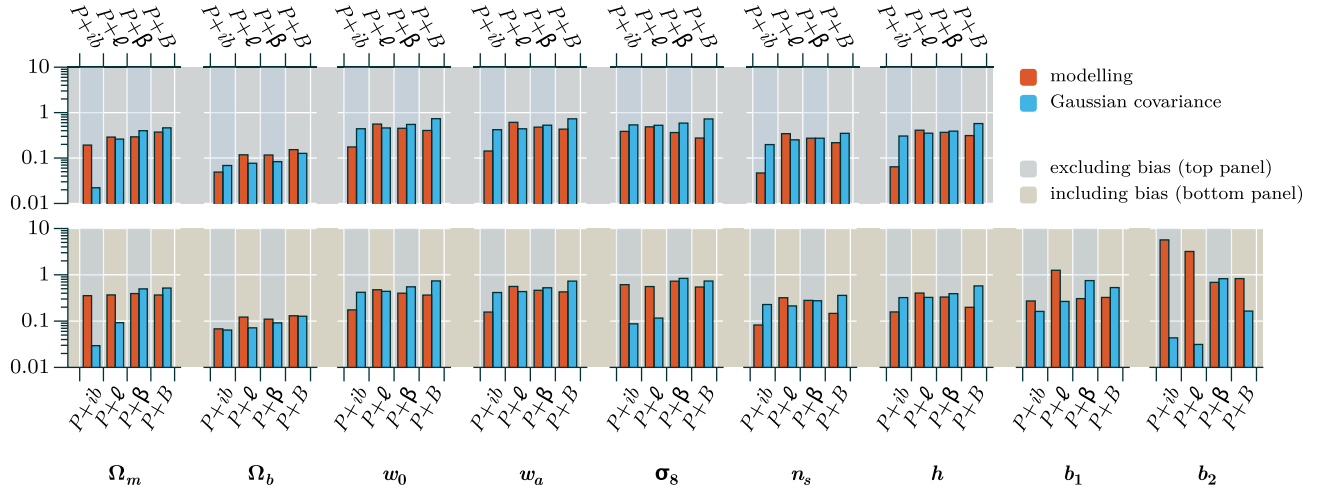


Figure 14. Fractional difference in predicted uncertainties induced by theoretical modelling of derivatives (orange) or by using a Gaussian approximation to the covariance (blue). (See the text for details of how the fractional differences are defined.)

7.4.2 Sensitivity to theory error

Next, we study the variation in forecasts for the Fourier bispectrum and its proxies when we change the model used to compute $\partial\mu/\partial\theta_\alpha$. To understand the sophistication required to obtain accurate models, we will need to understand which of these statistics (if any) are especially sensitive or immune to theoretical mis-modelling. We measure this dependence by a *sensitivity factor*, which we define to be the ratio between the largest and smallest forecast uncertainties taken over the models of Section 3. A sensitivity factor close to unity indicates that a forecast uncertainty depends only weakly on the choice of theoretical model, while a large value indicates that the model has a strong influence on the final outcome.

We plot these sensitivity factors in Fig. 13, computed with inclusion of all bias parameters and using non-Gaussian covariances. Therefore the sensitivity factor solely reflects the variation in uncertainty produced by different choices for theoretical model. We conclude that there is no single measure of 3-point correlations that consistently yields the largest or smallest sensitivity to variations in modelling. Therefore, there is apparently no single combination $P + X$ that should be preferred to minimize the effect of theory errors on inferred parameter constraints.

7.4.3 Ranking by constraining power

Neither of these criteria provides a rationale to prefer a choice of theoretical model. Nevertheless, we do find some general trends. Irrespective of theoretical model, we find the largest reductions in parameter uncertainties when the bias is constrained simultaneously with the cosmological parameters. Also, the Fourier bispectrum and modal bispectrum consistently offer the most significant improvements compared to P -only measurements, with very similar predicted uncertainties. The line correlation function achieves moderate improvement compared to P -only, while the integrated bispectrum has very weak constraining power – at least for the parameter set we consider. We conclude that $P + B$ or $P + \beta$ should be preferred for constraints on Λ CDM parameters, with $P + \beta$ offering similar information at reduced computation cost as discussed at the end of Section 7.3.

7.4.4 Relative importance of modelling and non-Gaussian covariance

Finally, we consider the relative importance of non-Gaussian covariance and theoretical modelling for obtaining quantitatively accurate forecasts. In Fig. 14 we show the fractional difference in Fisher forecasts induced by variation of theoretical model (orange bars) and the use of the Gaussian approximation (blue bars). To quantify the significance of theoretical modelling, we plot $\max(|\sigma_{NG,i}/\sigma_{NG}(\text{sim}) - 1|)$, where $i \in \{\text{tree, 1-loop, halo}\}$. Therefore larger orange bars reflect more significant deviation from the simulated forecast due to theoretical uncertainty. Meanwhile we quantify the role of the covariance matrix by plotting $|\sigma_G(\text{sim})/\sigma_{NG}(\text{sim}) - 1|$, so increasing blue bars show that the Gaussian approximation generates more significant errors in the forecast.

Table 4. Percent improvement of unmarginalized constraints using $P + B$ compared to P only at $z = 0$.

Ω_M	Ω_B	w_0	w_a	σ_8	n_s	h	b_1	b_2
12.9 per cent	19.4 per cent	26.0 per cent	27.0 per cent	26.4 per cent	15.1 per cent	15.6 per cent	42.4 per cent	43.4 per cent

Fig. 14 shows that the impact of theoretical uncertainty for $P + \beta$ and $P + B$ is generally less significant than neglect of non-Gaussian covariance, whether or not we marginalize over the bias. In contrast, for $P + \ell$ the effect of modelling nearly always dominates because of the difficulties with the halo model discussed in Section 5.2. For $P + ib$ the non-Gaussian covariance plays an important role if the bias parameters are not included, but theoretical modelling dominates when they are.

On balance, these results indicate that our forecasts are slightly less sensitive to theory error than to the approximation of Gaussian covariance. This could be because the inverse covariance weighting suppresses contributions from the non-linear regime where the theoretical predictions are most discrepant. But the difference is not large: the average variation in our predicted uncertainties from $P + B$ and $P + \beta$ due to theory modelling is 36 per cent, whereas the variation due to Gaussian covariances is 49 per cent. Therefore, we conclude that both issues must be addressed to obtain quantitatively accurate results.

7.5 Signal to noise as a proxy for the information content

It is now necessary to address the question of why the large discrepancy in uplift between the signal to noise of B and B_ϵ (equivalently β) observed in Section 6 did not translate into significant differences in the forecast for parameter uncertainties in Section 7.2.

Consider a vector of values S combining measures P and X of the 2- and 3-point correlation data, respectively, as defined below equation (54). For a given parameter θ the reduction in uncertainty compared to measurements from P alone can be estimated in the Fisher framework by

$$\frac{\mathbf{F}_\theta(S)}{\mathbf{F}_\theta(P)} = \sum_{i,j} \frac{\partial S_i}{\partial \theta} \mathbf{C}_{ij}^{-1} \frac{\partial S_j}{\partial \theta} \bigg/ \sum_{i,j} \frac{\partial P_i}{\partial \theta} (\mathbf{C}^P)_{ij}^{-1} \frac{\partial P_j}{\partial \theta}. \quad (63)$$

To avoid ambiguity, we use the notation \mathbf{C}^P to denote the covariance matrix of the power spectrum *only*. Meanwhile, the increase in signal to noise in the same scenario is given by

$$\frac{(S/\mathcal{N})_S^2}{(S/\mathcal{N})_P^2} = \sum_{i,j} S_i \mathbf{C}_{ij}^{-1} S_j \bigg/ \sum_{i,j} P_i (\mathbf{C}^P)_{ij}^{-1} P_j. \quad (64)$$

The uplift in signal to noise is often taken as an approximation to the reduction in parameter uncertainty, which avoids the need to compute $\partial S_i / \partial \theta$. As we have seen in Section 5.2, these derivatives can be rather fragile and are susceptible to significant errors caused by theory mis-modelling. Unfortunately, when applied to $S = P + B$ and $S = P + \beta$ our analysis demonstrates that the ratios $\mathbf{F}_\theta(P + B) / \mathbf{F}_\theta(P)$ and $\mathbf{F}_\theta(P + B_\epsilon) / \mathbf{F}_\theta(P)$ are nearly equal, whereas the same ratios constructed using S/\mathcal{N} are very discrepant. Therefore we must conclude that improvements in signal to noise cannot always be interpreted as a predictor of the improvement in Fisher information.

7.5.1 Invariance of the Fisher matrix

First consider the Fisher matrix. Suppose we perform a redefinition so that $S_i \rightarrow S'_i = S'_i(S_j)$, where S'_i may be an arbitrary non-linear function of the original measurements. For example, the transformation from B to B_ϵ is of this type. The derivative $\partial S_i / \partial \theta_\alpha$ transforms ‘contravariantly’ on its index i , in the sense $\partial S'_i / \partial \theta_\alpha = \sum_m (\partial S'_i / \partial S_m) (\partial S_m / \partial \theta_\alpha)$. Meanwhile, the covariance matrix becomes

$$\mathbf{C}_{ij}^S \rightarrow \mathbf{C}_{ij}^{S'} = \langle (S'_i - \bar{S}'_i)(S'_j - \bar{S}'_j) \rangle = \sum_{m,n} \frac{\partial S'_i}{\partial S_m} \frac{\partial S'_j}{\partial S_n} \mathbf{C}_{mn}^S + \dots, \quad (65)$$

where ‘...’ denotes terms involving higher order correlations that we have not written explicitly. Provided these are small compared to the \mathbf{C}_{mn}^S term, equation (65) shows that the covariance matrix also transforms ‘contravariantly’, and therefore that its inverse transforms ‘covariantly’. Subject to these approximations we conclude that the Fisher matrix should be roughly *invariant*. This agrees with our observation that $\mathbf{F}_\theta(P + B)$ and $\mathbf{F}_\theta(P + B_\epsilon)$ are nearly equal, demonstrated numerically in Table 3.

Now consider the signal to noise. Since S_i has neither a co- or contravariant transformation law, the combination $\sum_{i,j} S_i \mathbf{C}_{ij}^{-1} S_j$ appearing in the signal to noise will typically *not* be invariant. Therefore different choices S_i and S'_i may yield inequivalent results for S/\mathcal{N} . For example, we have verified that using $P + \ln B$ predicts a significant increase in the signal-to-noise compared to $P + B$, whereas their Fisher matrices continue to agree. In Table 4 we summarize the improvement in unmarginalized constraints from the addition of B or B_ϵ . This demonstrates that empirically the increase in signal-to-noise from B_ϵ provides a more accurate estimate of the Fisher information than B . This property holds for both proxies of B_ϵ , namely the modal bispectrum, and the line correlation function.

7.5.2 Gaussian limit

This outcome is not inconsistent with the result that B and B_ϵ show an equivalent uplift in signal to noise in the Gaussian approximation. In this case the covariance matrix for B_ϵ is $\mathbf{C}_{ij}^{B_\epsilon} = \mathbf{N}\mathbf{1}_{ij}$, where the constant \mathbf{N} takes the values 1, 2 or 6 for scalene, isosceles and equilateral configurations, respectively, as described in Section 3. In the same approximation the covariance matrix for the Fourier bispectrum is $\mathbf{C}_{ij}^B = \mathbf{N}P(k_{i_1})P(k_{i_2})P(k_{i_3})\mathbf{1}_{ij}$. Therefore we conclude that the signal to noise for B and B_ϵ is identically equal as

$$B_i(\mathbf{C}_{ij}^B)^{-1}B_j = B_{\epsilon i}(\mathbf{C}_{ij}^{B_\epsilon})^{-1}B_{\epsilon j} = \frac{1}{\mathbf{N}} \frac{B_i^2 \mathbf{1}_{ij}}{P(k_{i_1})P(k_{i_2})P(k_{i_3})}. \quad (66)$$

In the Gaussian approximation the power spectrum is an independent source of information, which explains the agreement. However, once off-diagonal contributions in the covariance matrix are included, B and P are no longer independent and non-linear combinations may give very different results for the signal to noise.

8 DISCUSSION

8.1 Compression and efficiency of the Fourier bispectrum proxies

In an ideal survey aiming to measure the Fourier bispectrum, we should clearly choose a bin width Δk that is sufficiently small to reproduce all small-scale features of interest. However, because the number of Fourier configurations in a volume with mode cut-off k_{\max} scales as $\sim(k_{\max}/\Delta k)^3$ this task will quickly become computationally expensive. And, as we have emphasized several times, a more serious problem is that we must estimate and invert the covariance matrix for all these measurements. This requires us to perform at least as many N -body simulations as the number of configurations that we retain.

In this section we consider how well this large number of Fourier configurations can be compressed by the proxies described in Section 2. Suppose that available resources limit the number of simulations that can be performed in such a way that we can estimate an accurate covariance matrix for ~ 30 bins of the Fourier bispectrum or one of its proxies, in combination with another 30 measurements of the power spectrum $P(k)$. Among the measures of 3-point correlations that we consider, is there a preferred choice that provides optimal constraints on our set of cosmological parameters? If so, this measure would provide the most successful compression of the full Fourier bispectrum into a manageable number of measurements.

8.1.1 Compression by reduction to ≤ 30 bins

To this end we combine the power spectrum bins with a single additional configuration from the Fourier bispectrum or one of its proxies, and compute the corresponding Fisher matrix (as in Section 7.1) using values for $\partial\boldsymbol{\mu}/\partial\theta_\alpha$ estimated from our simulation suite. The four left-hand panels of Fig. 15 show the reduction in predicted uncertainty – defined as the shrinkage of the error bar, $1 - \sigma_{P+X}/\sigma_P$ – for the representative parameters σ_8 (solid lines) and w_0 (dotted lines) for each of the possible bins. Using these reductions as a measure of the information stored in each bin, we conclude that most of the information carried by the Fourier bispectrum B is contained in small-scale triangles (towards larger triangle index). A similar conclusion applies for the line correlation function, for which significant reductions occur only for the first ~ 12 bins, corresponding to the range of scales $10\text{--}50 h^{-1}$ Mpc. This is reasonable, because the line correlation is constructed to give a negligible signal on large scales. Finally, while the modal decomposition exhibits some variability, smaller mode numbers typically provide larger gains. The integrated bispectrum shows consistently weak improvements over all bins.

Second, for each combination $P + X$ we identify a set of 30 bins for X that provide the largest improvements. Adding them cumulatively to the power spectrum, starting from the bin carrying most information, we obtain the plot on the right-hand side of Fig. 15. Both the line correlation function and the modal bispectrum converge rapidly to the maximal improvement available from the entire set of bins that we measure (this is 30 bins for ℓ and 50 modes for β ; see Table 2). For example, the line correlation is already within 2 per cent of the maximum after we have added ~ 2 bins, while only ~ 5 modes of β are required to arrive at a similar value for the modal bispectrum. In comparison, the Fourier bispectrum converges much more slowly to the maximum provided by the 95 bins that we measure. This is especially evident for σ_8 , for which the improvement from the Fourier bispectrum has not yet converged to its maximum value after the 30th bin. (For guidance, we mark this maximum value with black arrows on the plot.) However, it should be noted that our procedure to select the set of 30 bins is not optimal because it does not account for covariances between them. By analysing random subsets of the 95 possible bispectrum bins, we find that faster convergence is possible, giving up to ~ 90 per cent of the maximum reduction after 30 bins.

8.1.2 Compression by broadening bins

Rather than reducing the number of configurations by restriction to a subset, we might alternatively increase the width of each bin. The same volume of data would then be compressed into fewer measurements. To compare the performance of this strategy, we repeat the analysis described above for the Fourier bispectrum with a broader bin width $\Delta k = 12k_f$, which gives 34 rather than 95 Fourier configurations with $k_{\max} = 0.3 h \text{ Mpc}^{-1}$. We plot the corresponding cumulative reduction in uncertainty for σ_8 as star-shaped symbols in the right-hand panel of Fig. 15. After 30 bins the improvement is similar to that obtained from the modal bispectrum, with the same caution about rate of convergence due to correlation between bins. Therefore – rather surprisingly – in this case we find no clear preference for the bin width $\Delta k = 8k_f$ or

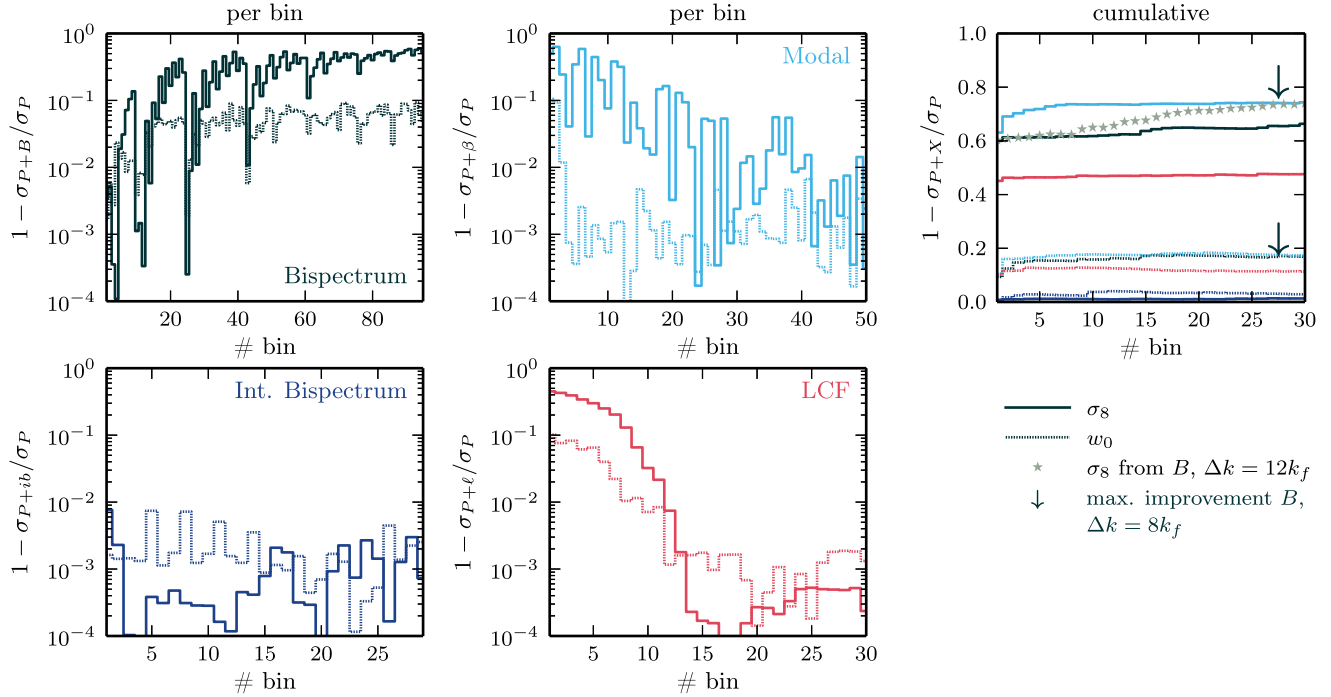


Figure 15. First two columns: decrease in forecast parameter uncertainty (improvement) from combining the power spectrum with a single bin of a 3-point correlation measure, compared to the power spectrum alone. The Fisher matrix was computed from the non-Gaussian covariance matrix and the measured parameter derivatives $\partial\mu/\partial\theta_a$. Solid (dotted) lines show σ_8 (w_0) with all other parameters (including bias) marginalized. Third column: cumulative improvement from adding the 30 best bins. Arrows indicate the maximal improvement obtained from the Fourier bispectrum with $\Delta k = 8k_f$, while stars show the uncertainty for σ_8 using Fourier bispectrum measurements with the larger bin width $\Delta k = 12k_f$.

$\Delta k = 12k_f$, except that $\Delta k = 8k_f$ is more computationally expensive, and it is more difficult to find an optimal subset of configurations. However, it is not clear whether this conclusion would survive in a more realistic analysis, where the signal can be noisy and demands finer binning. To explore these issues in detail would require a more comprehensive analysis.

8.1.3 Results

This analysis agrees with the conclusions of Sections 7.3 and 7.4, and supports the modal bispectrum as a good choice of proxy for 3-point correlation data. In addition to the advantages discussed in previous sections, it requires the fewest bins and loses almost no information.

These results could be modified in cases where it is possible to compute a covariance matrix for $\gg 30$ configurations of the Fourier bispectrum, as done (for example) by Gil-Marín et al. (2017). However, the mock catalogues used to produce such covariance matrices are often generated using perturbation theory and therefore are likely to be inaccurate on small scales. We expect that it is a better strategy to use fewer bins and obtain high-quality measurements of the covariance matrix from catalogues generated using full N -body simulations. The significant benefit of the modal decomposition is that it facilitates construction of the smallest set of bins that still carry a majority of the information.

Finally, although the line correlation function provides weaker improvements than either the Fourier bispectrum or modal bispectrum, it has the advantage that it clearly separates the scales carrying useful information from those that do not – all bins with $r \gtrsim 50 h^{-1}$ Mpc have negligible impact. It is also possible that the performance of the line correlation function could be improved by relaxing the condition of strict collinearity, which would increase the range of Fourier configurations it is able to aggregate.

8.2 Shot noise

Galaxies are discrete, point-like tracers of the underlying matter fluctuations, and therefore samples of their abundance are affected by shot noise. This noise is expected to impact higher-order statistics more significantly than the power spectrum (Sefusatti & Scoccimarro 2005; Chan & Blot 2016). Up to this point our analysis has implicitly used the low effective shot noise provided by our simulations, and therefore there is some concern that our forecasts will degrade with larger, more realistic noise. In this section we perform an approximate analysis of this degradation and quantify its effect on our predicted parameter uncertainties.

Assuming Poisson statistics, we may correct for shot-noise contributions to the observed discrete power spectrum \hat{P}^{disc} and bispectrum \hat{B}^{disc} by subtraction (Peebles 1980; Matarrese et al. 1997),

$$\hat{P}(k) = \hat{P}^{\text{disc}}(k) - \frac{1}{\bar{n}}, \quad (67a)$$

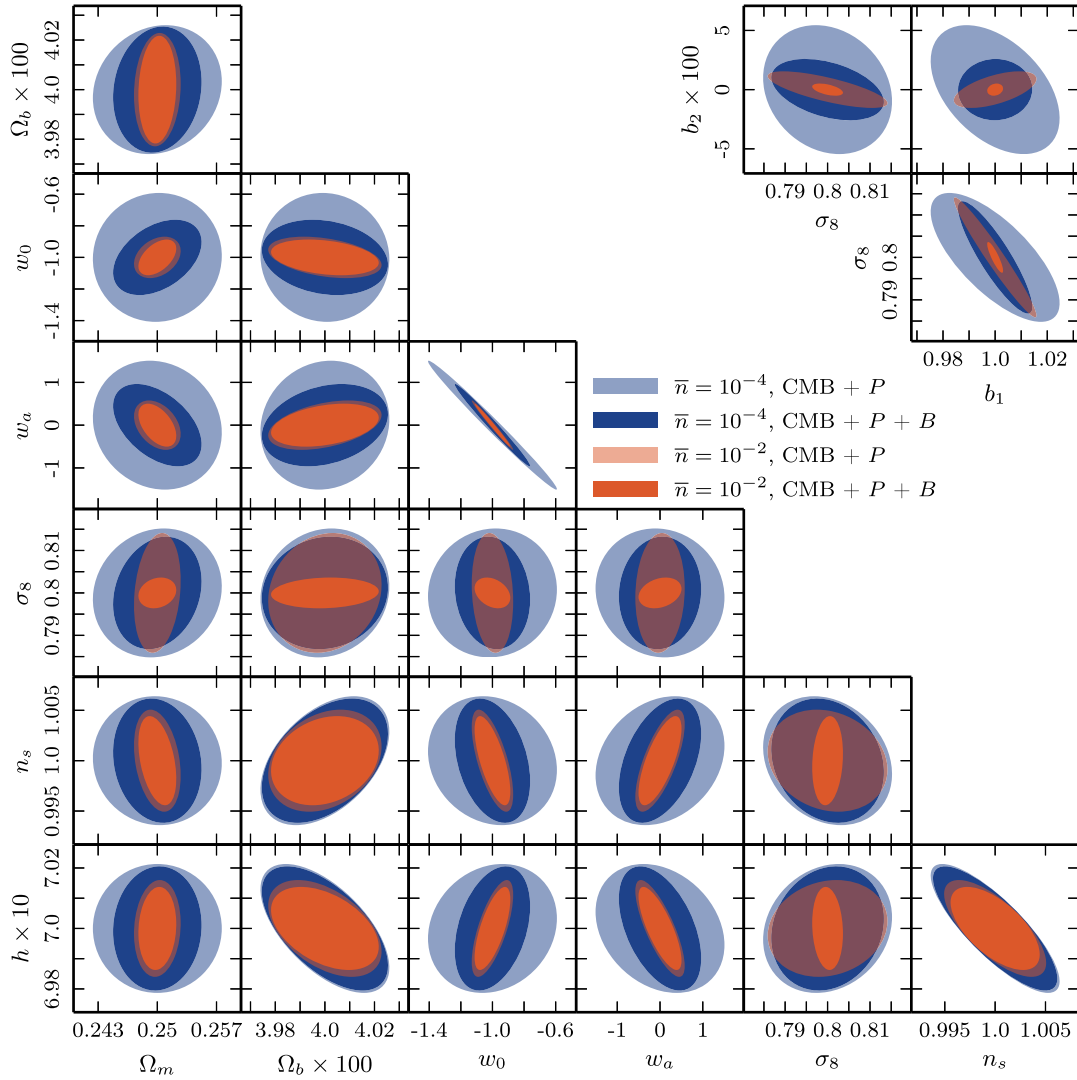


Figure 16. Comparison of the Fisher forecasts with shot noise corresponding to $\bar{n}_1 = 10^{-2} h^3 \text{Mpc}^{-3}$ (orange) and $\bar{n}_2 = 10^{-4} h^3 \text{Mpc}^{-3}$ (blue). The pale ellipses correspond to uncertainties using the power spectrum only, while the dark ellipses show the predicted uncertainty when 3-point correlation information is included.

$$\hat{B}(k_1, k_2, k_3) = \hat{B}^{\text{disc}}(k_1, k_2, k_3) - \frac{1}{\bar{n}} [\hat{P}(k_1) + \hat{P}(k_2) + \hat{P}(k_3)] - \frac{1}{\bar{n}^2}. \quad (67b)$$

Here, \bar{n} is the average number density of the discrete tracers. We use the upper and lower limits $\bar{n}_1 = 10^{-2} h^3 \text{Mpc}^{-3}$ and $\bar{n}_2 = 10^{-4} h^3 \text{Mpc}^{-3}$ to represent optimistic and pessimistic levels of shot noise for upcoming galaxy surveys. To measure \hat{P}^{disc} and \hat{B}^{disc} , we downsample the number of particles in our simulation suite by selecting random subsets matching the desired averaged density \bar{n} , and use this to compute corrected estimators \hat{P} and \hat{B} from equations (67a) and (67b). Although this downsampling procedure will not introduce exactly Poisson shot noise, we have checked that it is nearly Poisson by verifying that the corrected quantities agree with measurements made using the full set of particles to within a few percent. Strictly speaking, the covariance matrix of \hat{P} and \hat{B} obtained in this way is the matter covariance with Poisson shot noise, but for our fiducial biasing model we may interpret it as the covariance of the galaxy power spectrum and bispectrum with Poisson shot noise. We use this covariance, leaving the parameter derivatives unchanged from Section 7.2, to compute the Fisher matrices.

We plot forecasts using the fiducial number densities \bar{n}_1 and \bar{n}_2 in Fig. 16, with orange ellipses corresponding to the lower noise level (higher number density) and blue ellipses corresponding to the higher noise level (lower number density). The orange ellipses show good agreement with the forecasts for the idealized scenario of Section 7.2, indicating that relatively little degradation occurs. However, it is unlikely that such high number densities will be attained in the near future. By contrast the blue ellipses represent a conservative view of what should be possible.

If shot noise degrades the signal from 3-point correlations more strongly than for 2-point correlations, then the fractional improvement from its inclusion should be smaller for low \bar{n} . In terms of Fig. 16 this means that the difference between the light and dark blue ellipses should be smaller than the difference between the light and dark orange ellipses. This effect is visible for some parameters, such as σ_8 . However, in

the case of Ω_m , w_0 and w_a the fractional improvement from inclusion of 3-point correlation data is *larger* at lower \bar{n} . The effect for w_0 and w_a is particularly striking. Using all particles in our simulations, the addition of B data decreased measurement uncertainties by 16 per cent and 15 per cent, respectively (see Table 3). With $\bar{n} = 10^{-4} h^3 \text{Mpc}^{-3}$ we find improvements of 41 per cent and 36 per cent. There are (at least) two possible explanations for this effect. First, while shot noise puts a stricter limit on the range of k -modes that are useful to recover cosmological information from 3-point correlations than from 2-point correlations, it also tends to render the respective covariance matrices more Gaussian, by relatively decreasing the correlation in the off-diagonal terms. That means different bins are less correlated and as the bispectrum allows for many more configurations than the power spectrum, this might enhance the relative importance of 3-point information. Alternatively, recovery of cosmological information in the presence of shot noise might significantly depend on cross-covariances between measurements. These cross-covariances themselves depend on the shot noise and can partially subtract its effect. By explicitly excluding the cross-covariance between \hat{P} and \hat{B} , we find that the improvement in uncertainty from inclusion of the 3-point information drops slightly from 10 per cent (9 per cent) to 8 per cent (7 per cent) for w_0 (w_a) as shot noise increases, which strongly suggests that this second possibility is the driving force in the improvements of w_0 and w_a .

9 CONCLUSIONS

As large-scale structure surveys grow in size and sophistication, the rapidly approaching cosmic variance limit on 2-point statistics encourages us to look to higher-order correlations, such as the 3-point function, as a new source of information. Previously, Sefusatti et al. (2006) suggested that considerable additional constraining power could be achieved by combining the power spectrum and bispectrum. On the other hand, the signal-to-noise analysis in Chan & Blot (2016) concluded that the bispectrum yields no more than modest improvements. Our results show that there is a significant benefit from inclusion of 3-point correlation data, but its benefits must be balanced against the challenges it brings.

In this paper, we focus on two particular challenges: (1) The number of measurable configurations of the Fourier bispectrum is generally very large unless one coarse-grains the data. We have investigated whether the *modal bispectrum*, *line correlation function* and *integrated bispectrum* can act as ‘proxies’ for the Fourier bispectrum, compressing its information into fewer configurations without unacceptable information loss. (2) Bispectrum observations are difficult to model to the same accuracy as the power spectrum. Errors in clustering predictions from theoretical models, in addition to assumptions about covariances and noise properties, generally propagate into inaccurate error bars or a bias on inferred parameters. We have quantified how our forecasts are influenced by both the assumption of Gaussian covariance and theoretical errors.

To do so, we have measured the power spectrum, Fourier bispectrum and each of its proxies from a suite of 200 dark matter N -body simulations at redshifts $z = 0$, $z = 0.52$ and $z = 1$ to obtain fully non-Gaussian covariances and cross-covariances. We measure the dependence of each measurement on the cosmological parameters $\{\Omega_m, \Omega_b, w_0, w_a, \sigma_8, n_s, h\}$ using additional simulations displaced from our fiducial model. We assume a local Lagrangian biasing scheme that includes two bias parameters, $\{b_1, b_2\}$. Using all these components, in combination with theoretical predictions for each proxy from tree-level and 1-loop SPT and the halo model, we have conducted a signal-to-noise analysis and implemented the Fisher forecasting method for an idealized survey scenario. Our main results on the constraining power and future viability of each measure of 3-point correlations are as follows.

9.1 Comparison of 3-point correlation measures

Section 7.2 presents our main results. Our forecasts show that inclusion of the Fourier bispectrum offers significant improvements over the power spectrum alone, with $\mathcal{O}(10\text{--}30)$ per cent improvement on cosmological parameter constraints, and up to $\mathcal{O}(80)$ per cent improvement when it is used to break degeneracies with the bias parameters. The *modal bispectrum* offers an attractive alternative, achieving equivalent constraints with as few as 10 modes. However, up to 50 modes may be necessary to reconstruct the Fourier bispectrum to within $\lesssim 10$ per cent accuracy on individual triangle configurations. The *line correlation function* appears to be slightly less optimal, although a future extension to sample more Fourier configurations by relaxing the requirement of strict collinearity may improve its performance. The *integrated bispectrum* offers little constraining power for our set of cosmological parameters. It is sensitive to highly squeezed triangles, whereas the gravitational bispectrum peaks on equilateral triangles. This property of *ib* is a disadvantage for our purposes, but may be an advantage if one is interested in studying squeezed-mode primordial non-Gaussianity with minimal degeneracies.

9.2 Data compression

In Section 8.1, we explored how the total constraining power of each measure is distributed over the total number of data bins. While the Fourier bispectrum and modal bispectrum give nearly equivalent parameter constraints when ~ 30 bins are used, the modal method converges to its full constraining power with a smaller subset of bins. We conclude that the modal bispectrum provides more efficient access to the information carried by 3-point correlations.

We note that more realistic survey scenarios – for example, accounting for noisy data – may require finer binning. Increasing the binning resolution of the Fourier bispectrum by a factor of n in each k -dimension corresponds to a factor $\mathcal{O}(n^3)$ increase in configurations. The number of simulations required to accurately capture their covariance would increase similarly. If the number of modal coefficients required to capture fine features of the bispectrum does not grow so dramatically, it is possible that the modal bispectrum could accumulate an even larger advantage compared to the Fourier bispectrum.

9.3 Signal-to-noise ratio as a measure of information content

In Sections 6, 7.2 and 7.5, we argue that the use of the signal-to-noise ratio to predict the constraining power of 3-point correlation data can be misleading. We show that the bispectrum and phase bispectrum – which is probed by the modal bispectrum – give significantly different signal-to-noise ratios, but still yield nearly identical forecasts. As we describe in Section 7.5, for the scenarios considered in this paper, the improvement shown by these forecasts is empirically better predicted by the signal-to-noise ratio of the phase bispectrum B_e than the Fourier bispectrum B . The $\sim\mathcal{O}(30\text{ per cent})$ uplift in signal to noise from the phase bispectrum translates to the same improvement in cosmological parameter constraints, except for those where degeneracies play a significant role. For a general parameter set and a given measure of the 3-point correlations, the signal to noise will not typically give an accurate estimate of its constraining power.

9.4 Impact of non-Gaussian covariances

Accounting for non-Gaussian covariance is essential for optimally constraining cosmological parameters. In Section 7.3 we showed that the Fourier bispectrum estimator is particularly sensitive to the covariance: our predicted uncertainties may be nearly a factor of 4 too small if the Gaussian approximation is used. At the same time, we find that the non-Gaussian cross-covariance between the power spectrum and the Fourier bispectrum or its proxies generally results in parameter errors that are $\mathcal{O}(10\text{ per cent})$ *smaller* than if cross-covariances are ignored.

9.5 Impact of theoretical modelling uncertainties

Our results in Section 7.4 indicate that the impact of theory errors on our predicted uncertainties is smaller than the impact of assuming Gaussian covariance, although both approximations change the forecasts by $\sim 30\text{ per cent}$ to 50 per cent on average. In this paper we measure the effect of theoretical uncertainty by comparing forecasts using SPT and the halo model to forecasts derived purely from N -body measurements. Our approach differs from that of Baldauf et al. (2016) and Welling et al. (2016), who incorporated estimates of the theory error into their Fisher forecasts by taking the error in each data bin to be the sum of statistical and theoretical errors.

9.6 Impact of shot noise

To assess the impact of shot noise, in Section 8.2 we downsample our simulation suite to averaged number densities of $\bar{n} = 10^{-2} h^3 \text{ Mpc}^{-3}$ and $10^{-4} h^3 \text{ Mpc}^{-3}$, and compute forecasts using non-Gaussian covariance matrices that include low and high levels of Poisson shot noise. Contrary to naïve expectations, we find that the addition of 3-point correlation information can become *more* significant at high levels of shot noise owing to the non-trivial dependence of the cross-covariance on \bar{n} . This appears most significant for the dark energy parameters w_0 and w_a , and suggests that 3-point correlation information may be crucial to distinguish between dark energy models. More generally, our result implies that 3-point correlation measurements may yield significant additional constraining power even when shot noise levels are high.

To make robust inferences with 3-point correlation information, future surveys will require refinement of the methods we have considered here. For example, while we have demonstrated that the modal decomposition provides efficient data compression of the matter bispectrum in an idealized survey, it will be important to verify that this remains true when halo distributions, redshift-space distortions and the complex noise properties of realistic surveys are introduced. We have emphasized the importance of including non-Gaussian covariances and theory uncertainties in our forecasts. Realistic analyses will likely require more efficient ways to obtain covariances, and a consistent approach to inclusion of theory errors in software pipelines. Achieving each of these aims will be important milestones ahead of upcoming surveys of a large-scale structure.

ACKNOWLEDGEMENTS

The work reported in this paper has been supported by the European Research Council under the European Union’s Seventh Framework Programme (FP/2007–2013) and ERC Grant Agreement No. 308082 (JB, DR, DS). This work was supported by the Science and Technology Facilities Council [grant numbers ST/L000652/1, ST/P000525/1] (DS, RES). AE acknowledges support from the UK Science and Technology Facilities Council via Research Training Grant (grant number ST/M503836/1), and thanks Roman Scoccimarro and the Physics Department of New York University for hospitality during the final phases of this project. DR acknowledges useful conversations on the normalization of the modal decomposition with Hemant Shukla. JB would like to thank Benjamin Joachimi for useful discussions. The authors also thank the referee for an illuminating deliberation regarding the shot noise results of Section 8.2, which has been recounted therein.

Data availability statement. To assist those wishing to replicate or extend our results, we have made available measurements of the power spectrum, bispectrum, integrated bispectrum, line correlation function and modal bispectrum coefficients that have been extracted from our N -body simulation suite.

REFERENCES

- Ade P. A. R. et al., 2014, *A&A*, 571, A24
- Anderson T. W., 2003, *An Introduction to Multivariate Statistical Analysis*. Wiley, New York
- Assassi V., Baumann D., Green D., Zaldarriaga M., 2014, *JCAP*, 1408, 056
- Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, *Phys. Rev. D*, 86, 083540
- Baldauf T., Mirbabayi M., Simonovi M., Zaldarriaga M., 2016, preprint ([arXiv:1602.00674](https://arxiv.org/abs/1602.00674))
- Bennett C. et al., 2003, *ApJS*, 148, 97
- Bernardeau F., Colombi S., Gaztanaga E., Scoccimarro R., 2002, *Phys. Rept.*, 367, 1
- Blot L., Corasaniti P. S., Amendola L., Kitching T. D., 2016, *MNRAS*, 458, 4462
- Carron J., Szapudi I., 2013, *MNRAS*, 434, 2961
- Catelan P., Porciani C., Kamionkowski M., 2000, *MNRAS*, 318, 39
- Chan K. C., Blot L., 2016, preprint ([arXiv:1610.06585](https://arxiv.org/abs/1610.06585))
- Chan K. C., Scoccimarro R., Sheth R. K., 2012, *Phys. Rev. D*, 85, 083509
- Chevallier M., Polarski D., 2001, *Int. J. Mod. Phys. D*, 10, 213
- Chiang C.-T., 2015, PhD thesis, Munich Univ
- Chiang C.-T., Wagner C., Schmidt F., Komatsu E., 2014, *JCAP*, 1405, 048
- Chiang C.-T., Wagner C., Sánchez A. G., Schmidt F., Komatsu E., 2015, *JCAP*, 1509, 028
- Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
- Crocce M., Pueblas S., Scoccimarro R., 2006, *MNRAS*, 373, 369
- Efstathiou G., 1999, *MNRAS*, 310, 842
- Eggemeier A., Smith R. E., 2017, *MNRAS*, 466, 2496
- Eggemeier A., Battefeld T., Smith R. E., Niemeyer J., 2015, *MNRAS*, 453, 797
- Fergusson J. R., Shellard E. P. S., 2009, *Phys. Rev. D*, 80, 043510
- Fergusson J. R., Liguori M., Shellard E. P. S., 2010, *Phys. Rev. D*, 82, 023502
- Fergusson J. R., Regan D. M., Shellard E. P. S., 2012, *Phys. Rev. D*, 86, 063511
- Fry J. N., 1984, *ApJ*, 279, 499
- Fry J. N., 1994, *Phys. Rev. Lett.*, 73, 215
- Fry J. N., Gaztanaga E., 1993, *ApJ*, 413, 447
- Gil-Marín H., Noreña J., Verde L., Percival W. J., Wagner C., Manera M., Schneider D. P., 2015, *MNRAS*, 451, 539
- Gil-Marín H., Percival W. J., Verde L., Brownstein J. R., Chuang C.-H., Kitaura F.-S., Rodríguez-Torres S. A., Olmstead M. D., 2017, *MNRAS*, 465, 1757
- Goroff M. H., Grinstein B., Rey S. J., Wise M. B., 1986, *ApJ*, 311, 6
- Hahn T., 2016, *Comput. Phys. Commun.*, 207, 341
- Hartlap J., Simon P., Schneider P., 2006, *A&A*, 464, 399
- Joachimi B., Shi X., Schneider P., 2009, *A&A*, 508, 1193
- Kaiser N., 1984, *ApJ*, 284, L9
- Kayo I., Takada M., Jain B., 2013, *MNRAS*, 429, 344
- Klypin A., Prada F., 2017, preprint ([arXiv:1701.05690](https://arxiv.org/abs/1701.05690))
- Lazanu A., Giannantonio T., Schmittfull M., Shellard E. P. S., 2016, *Phys. Rev. D*, 93, 083517
- Lokas E. L., Juszkiewicz R., Bouchet F. R., Hivon E., 1996, *ApJ*, 467, 1
- Makino N., Sasaki M., Suto Y., 1992, *Phys. Rev. D*, 46, 585
- Marín F. A. et al., 2013, *MNRAS*, 432, 2654
- Matarrese S., Verde L., Heavens A. F., 1997, *MNRAS*, 290, 651
- McDonald P., Roy A., 2009, *JCAP*, 0908, 020
- Mead A., Peacock J., Heymans C., Joudaki S., Heavens A., 2015, *MNRAS*, 454, 1958
- Mead A., Heymans C., Lombriser L., Peacock J., Steele O., Winther H., 2016, *MNRAS*, 459, 1468
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Neyrinck M. C., Szapudi I., Szalay A. S., 2009, *ApJ*, 698, L90
- Obreschkow D., Power C., Bruderer M., Bonvin C., 2013, *ApJ*, 762, 115
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton, NJ
- Regan D. M., Shellard E. P. S., Fergusson J. R., 2010, *Phys. Rev. D*, 82, 023520
- Regan D. M., Schmittfull M. M., Shellard E. P. S., Fergusson J. R., 2012, *Phys. Rev. D*, 86, 123524
- Saito S., Baldauf T., Vlah Z., Seljak U., Okumura T., McDonald P., 2014, *Phys. Rev. D*, 90, 123522
- Sato M., Nishimichi T., 2013, *Phys. Rev. D*, 87, 123538
- Schmittfull M. M., Regan D. M., Shellard E. P. S., 2013, *Phys. Rev. D*, 88, 063512
- Scoccimarro R., 2000, *ApJ*, 544, 597
- Scoccimarro R., 2015, *Phys. Rev. D*, 92, 083532
- Scoccimarro R., Frieman J., 1996, *ApJ*, 473, 620
- Scoccimarro R., Sheth R. K., 2002, *MNRAS*, 329, 629
- Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998, *ApJ*, 496, 586
- Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, *ApJ*, 546, 20
- Scoccimarro R., Sefusatti E., Zaldarriaga M., 2004, *Phys. Rev. D*, 69, 103513
- Sefusatti E., 2005, PhD thesis, New York Univ., New York, USA
- Sefusatti E., Komatsu E., 2007, *Phys. Rev. D*, 76, 083004
- Sefusatti E., Scoccimarro R., 2005, *Phys. Rev. D*, 71, 063001
- Sefusatti E., Crocce M., Pueblas S., Scoccimarro R., 2006, *Phys. Rev. D*, 74, 023522
- Sefusatti E., Crocce M., Desjacques V., 2012, *MNRAS*, 425, 2903

- Seljak U., 2000, MNRAS, 318, 203
 Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437
 Sellentin E., Heavens A. F., 2016, MNRAS, 456, L132
 Seo H.-J., Sato M., Dodelson S., Jain B., Takada M., 2011, ApJ, 729, L11
 Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
 Silk J., 2017, Proc. 14th International Symposium on Nuclei in the Cosmos (NIC-XIV), Vol. 14, 010101 (Doi:10.7566/JPSCP.14.010101)
 Slepian Z. et al., 2015, MNRAS, preprint (arXiv:1512.02231)
 Slepian Z. et al., 2017, MNRAS, 469, 1738
 Smith R. E., 2009, MNRAS, 400, 851
 Smith R. E. et al., 2003, MNRAS, 341, 1311
 Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys. Rev. D, 75, 063512
 Smith R. E., Reed D. S., Potter D., Marian L., Crocce M., Moore B., 2014, MNRAS, 440, 249
 Springel V., 2005, MNRAS, 364, 1105
 Takada M., Bridle S., 2007, New J. Phys., 9, 446
 Takahashi R. et al., 2009, ApJ, 700, 479
 Takahashi R. et al., 2011, ApJ, 726, 7
 Tegmark M., Taylor A., Heavens A., 1997, ApJ, 480, 22
 Tellarini M., Ross A. J., Tasinato G., Wands D., 2016, JCAP, 1606, 014
 Welling Y., van der Woude D., Pajer E., 2016, JCAP, 1608, 044
 Wolstenhulme R., Bonvin C., Obreschkow D., 2015, ApJ, 804, 132

APPENDIX A: CONSTRUCTION OF THE MODAL DECOMPOSITION

A1 Construction of the Q basis

The goal of the modal decomposition is to write the estimated bispectrum in the form

$$w(k_1, k_2, k_3) \hat{B}(k_1, k_2, k_3) = \sum_{n=0}^{n_{\max}-1} \hat{\beta}_n^Q Q_n(k_1, k_2, k_3), \quad (\text{A1})$$

where $w(k_1, k_2, k_3)$ is the arbitrary weighting function (24), and Q_n represents basis modes with coefficients β_n^Q . The Q_n then contains all the information about the bispectrum. They should span the possible functions on wavenumbers k_i that satisfy the triangle condition, $\sum_i k_i \geq 2 \max\{k_1, k_2, k_3\}$ (denoted by \mathcal{V} in the main text) but are otherwise arbitrary. For our concrete numerical results, we choose a basis built out of 1-dimensional polynomials $q_p(x)$ that are orthonormal within \mathcal{V} (Fergusson et al. 2010). More precisely, in a unit box, we define the integral $\mathcal{T}[f] = \int_{\mathcal{V}} f(x) dx dy dz$, where x, y, z satisfy the triangle condition within the box $x, y, z \in [0, 1]$. Evaluating the y and z integrals, one finds that $\mathcal{T}[f] = 0.5 \int_0^1 f(x) x(4-3x) dx$. This allows one to define an inner product, $\langle f, g \rangle \equiv \mathcal{T}[fg]$ [which is not equal to the inner product (27)] and set up a generating function for the 1-dimensional polynomials, q_n , using $w_n = \mathcal{T}[x^n]$, in the form of a secular determinant

$$q_n(x) = \frac{1}{\mathcal{N}} \begin{vmatrix} 1/2 & 7/24 & \dots & w_n \\ 7/24 & 1/5 & \dots & w_{n+1} \\ \dots & \dots & \dots & \dots \\ w_{n-1} & w_n & \dots & w_{2n-1} \\ 1 & x & \dots & x^n \end{vmatrix} \quad (\text{A2})$$

where \mathcal{N} is chosen such that $\langle q_n, q_m \rangle = \mathbf{1}_{nm}$. The basis functions $Q_n(x, y, z)$ are defined as symmetric combinations of combinations of these 1-dimensional polynomials, in the form

$$Q_n(x, y, z) = \frac{1}{6} [q_r(x)q_s(y)q_t(z) + q_r(x)q_t(y)q_s(z) + \dots + q_t(x)q_s(y)q_r(z)] \equiv q_{[r}(x)q_s(y)q_{t]}(z), \quad (\text{A3})$$

with n representing the triple of indices $\{r, s, t\}$. After choosing an ordering of these triples, we can exchange n for a simpler integer label. For a particular realization with wavenumbers in the range k_{\min} and k_{\max} , we use the notation $Q_n(k_1, k_2, k_3)$ to represent $Q_n(x_1, x_2, x_3)$, where $x_i = (k_i - k_{\min})/(k_{\max} - k_{\min}) \in [0, 1]$.

A2 Calculation of the modal coefficients using the voxel method

In Section 4.7 we explained how equation (49) reduces estimation of the modal coefficients from simulation or data to a single 3-dimensional integral over a product of three Fourier transforms $\mathcal{M}_n(x)$. If the bispectrum is given analytically, however, we may instead use the simpler equation (27) and compute the inner product using a sum of volumes of all ‘voxels’ within a cubic grid with linear spacing along each axis (k_1, k_2, k_3) .

To calculate the volume of each voxel, we relabel the coordinates as (x, y, z) , rescaled so that $0 \leq x, y, z \leq 1$. We associate each of the 8 possible vertices of the voxel with a value p_1, \dots, p_8 , given by the product of Q_m and wB (or Q_m and Q_n in the case of $\langle Q_m | Q_n \rangle$) at that vertex. Finally, we define an interpolation function f by writing

$$f(x, y, z) = a_1 + a_2x + a_3y + a_4z + a_5xy + a_6xz + a_7yz + a_8xyz. \quad (\text{A4})$$

The coefficients a_i may be obtained analytically in terms of the p_i . We assign the volume of the voxel to be zero if fewer than four of its vertices satisfy the triangle condition, while if all eight vertices satisfy the triangle condition, its volume is

$$\int_{0 \leq x, y, z \leq 1} f(x, y, z) dx dy dz = \frac{1}{8} \sum_{i=1}^8 p_i, \quad (\text{A5})$$

as expected. For intermediate cases we write the volume in the form

$$\int_{\mathcal{C}} f(x, y, z) dx dy dz, \quad (\text{A6})$$

where \mathcal{C} indicates that only those points satisfying the triangle condition and forming a closed volume within the voxel should be included. In the case of 4 points there are 3 possible volumes given by

$$\mathcal{C}_a^{(4)} = \{x, y, z \mid x + 1 \leq y + z\}, \quad \mathcal{C}_b^{(4)} = \{x, y, z \mid y + 1 \leq x + z\}, \quad \mathcal{C}_c^{(4)} = \{x, y, z \mid z + 1 \leq x + y\}. \quad (\text{A7})$$

For 5 points the only possibility is that $x + y + z \geq 2\max\{x, y, z\}$, while for 6 and 7 points there are again 3 possibilities, given respectively by

$$\begin{aligned} \mathcal{C}_a^{(6)} &= \{x, y, z \mid x \leq y + z, y \leq x + z\}, & \mathcal{C}_b^{(6)} &= \{x, y, z \mid x \leq y + z, z \leq x + y\}, & \mathcal{C}_c^{(6)} &= \{x, y, z \mid y \leq x + z, z \leq x + y\}, \\ \mathcal{C}_a^{(7)} &= \{x, y, z \mid x \leq y + z\}, & \mathcal{C}_b^{(7)} &= \{x, y, z \mid y \leq x + z\}, & \mathcal{C}_c^{(7)} &= \{x, y, z \mid z \leq x + y\}. \end{aligned} \quad (\text{A8})$$

In each case the analytic form of the integral in terms of the vertex values p_i can be calculated easily. Computation of each integral using this voxel method is highly accurate and efficient.

This paper has been typeset from a \LaTeX file prepared by the author.